



# Spatial and Temporal Modeling of Heteroscedastic Volatility Behaviors in Social Science

著者	SATO TAKAKI
学位授与機関	Tohoku University
学位授与番号	11301甲第18407号
URL	<a href="http://hdl.handle.net/10097/00125703">http://hdl.handle.net/10097/00125703</a>

Spatial and temporal modeling of heteroscedastic  
volatility behaviors in social science

Takaki Sato

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>SPATIAL AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY MODELS</b>	<b>12</b>
2.1	Introduction . . . . .	12
2.2	Spatial Autoregressive Conditional Heteroskedasticity model . . .	14
2.3	Estimation . . . . .	16
2.3.1	Parameter estimation by the two step procedure . . . . .	16
2.3.2	Asymptotic results . . . . .	18
2.4	Empirical Analysis . . . . .	21
2.4.1	Simulation studies . . . . .	21
2.4.2	Land price data analysis . . . . .	22
2.5	Conclusion . . . . .	23
<b>3</b>	<b>Spatial GARCH Models</b>	<b>31</b>
3.1	Introduction . . . . .	31
3.2	Spatial GARCH models . . . . .	33
3.3	Estimation . . . . .	34
3.3.1	First step estimation . . . . .	34
3.3.2	Second step estimation . . . . .	37
3.4	Empirical analysis . . . . .	38
3.4.1	Simulation studies . . . . .	38
3.4.2	Land price data analysis . . . . .	38
3.5	Conclusion . . . . .	40
<b>4</b>	<b>SARAR-GARCH models</b>	<b>62</b>
4.1	Introduction . . . . .	62
4.2	Models . . . . .	64
4.3	Estimation . . . . .	65
4.4	Real data analysis . . . . .	66
4.4.1	Japanese market analysis . . . . .	66
4.4.2	U.S. market analysis . . . . .	68
4.5	Conclusion . . . . .	69

# Acknowledgments

First, I wish to express the deepest appreciation to my advisor, Professor Yasumasa Matsuda, for valuable guidance and encouragement. He is always willing to spend his time with me on research issues and giving suggestions. Without his patience and help, this dissertation would not have been possible.

Second, I am grateful to my dissertation committee members, Professor Nobuhiko Terui and Professor Tsukasa Ishigaki, for their valuable comments and encouragement. They always gave me helpful suggestions and stimulating comments. Moreover, I would also like to thank Professor Yoshihiro Yajima and Ryozi Miura for their valuable advice in seminars.

Finally, I have to express my gratitude to my family. They support my study long time.

# Chapter 1

## Introduction

This thesis is concerned with spatial and spatio-temporal extensions of time series volatility models. Volatility which is a variance and conditional variance in a model is one of the most important concepts in financial econometrics because it is used in widely areas such as risk management, option pricing and portfolio selection. However, volatility has some special features. The daily volatility is not directly observable from market data because we observe only one observation in a trading day. These unobservability makes it difficult to evaluate and forecast volatility. Moreover, financial market data often exhibits volatility clustering (i.e., volatility may be high for certain time periods and low for other periods). This means time-varying volatility is more common than constant volatility. Therefore, accurate modeling of time-varying volatility is important in financial econometrics.

The seminal work of Engle (1982) proposes autoregressive conditional heteroscedasticity (ARCH) models and the most important extension of the model is generalized ARCH (GARCH) models proposed by Bollerslev (1986). Let  $r_t$  be the log return of an asset at time index  $t$ . The GARCH (1, 1) model for  $r_t$  is defined by

$$\begin{aligned} r_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

where  $\sigma_t$  is volatility,  $\varepsilon_t$  is random variable belonging to an independent and identically distributed (i.i.d.) process with mean equal to 0 and variance equal to 1,  $\omega, \alpha$  and  $\beta$  are parameters. The positivity of  $\sigma_t$  is ensured by the following sufficient restrictions:  $\omega > 0, \alpha \geq 0$  and  $\beta \geq 0$ . The return  $r_t$  is uncorrelated but not independent and the dependence of  $r_t$  can be described by a linear combination of quadratic function of its lagged values and volatilities. After that, Many extended GARCH models have been proposed. For example, integrated GARCH models (Engle and Bollerslev (1986)), exponential GARCH models (Nelson (1991)), threshold GARCH models (Glosten, et al (1993)), GARCH in

the mean models, and GJR-GARCH models are proposed.

Univariate volatility models are generalized to multivariate cases in many ways. Financial assets tend to move together over time across both assets, therefore modeling a time-varying covariance matrix is important in many financial applications. Let  $\mathbf{r}_t = (r_{1,t}, \dots, r_{n,t})$  denote a  $n$ -dimensional vector process. As a natural extension of univariate models, processes are defined by,

$$\mathbf{r}_t = \boldsymbol{\Sigma}_t^{\frac{1}{2}} \boldsymbol{\varepsilon}_t,$$

where  $\boldsymbol{\Sigma}_t^{\frac{1}{2}}$  is a symmetric matrix square-root of a covariance matrix  $\boldsymbol{\Sigma}_t$ , such that  $\boldsymbol{\Sigma}_t^{\frac{1}{2}} \boldsymbol{\Sigma}_t^{\frac{1}{2}} = \boldsymbol{\Sigma}_t$  and  $\boldsymbol{\varepsilon}_t$  are standardized innovations with mean zero and a covariance matrix equal to the identity. The curse of dimensionality becomes a major obstacle for generalization because there are  $\frac{n(n+1)}{2}$  quantities in a covariance matrix for a  $n$ -dimensional time series. Thus, we attempt to give a conditional covariance matrix some simple structures to reduce the number of parameters. For example, exponentially weighted moving average models, constant conditional correlation models (Bollerslev (1990)), BEKK models (Engle and Kroner (1995)), orthogonal GARCH models (Alexander (2001)), dynamic conditional correlation models (Tse and Tsui (2002)), dynamic orthogonal component models, and factor GARCH models are proposed.

In this thesis, we apply spatial econometrics ideas to overcome the curse of dimensionality. Spatial econometrics models deal with cross-sectional correlation between observations. Therefore, spatial econometrics models also face the curse of dimensionality because there are  $n^2 - n$  relations that could arise. The solution to the problem is to impose structure on the spatial dependence relations. We use spatial weight matrices to express the dependence relations between observations and to reduce the number of parameters. Spatial weight matrices are calculated based on distance between observations.

The ideas of spatial econometrics have been applied to volatility models in recent years. Two main objectives of the applications are to reduce parameters in covariance matrices and to extend time series volatility models to spatial models. Caporin and Paruolo (2008) and Borovkova and Lophuua (2012) have applied the ideas of spatial econometrics to time series multivariate GARCH models from the former view point. On the other hand, Yan (2007) and Robinson (2009) have done spatial extensions of stochastic volatility models which are another kind of volatility models and Sato and Matsuda (2017) have extended time series ARCH models to spatial ARCH (S-ARCH) models from both view points.

We propose spatial ARCH (S-ARCH) models, spatial GARCH (S-GARCH) models and spatial autoregressive models with spatial autoregressive error and generalized autoregressive conditional heteroskedasticity processes (SARAR-GARCH) in this thesis. S-ARCH models and S-GARCH models are spatial extensions of ARCH and GARCH models for spatial data. SARAR-GARCH models are a spatio-temporal extension of ARCH models which are defined by spatial weight matrices based on financial distance.

In chapter 2, we propose a spatial extension of time series autoregressive conditional heteroskedasticity (ARCH) models to those for areal data. We call the spatially extended ARCH models as spatial ARCH (S-ARCH) models. Suppose we have areal data  $y_i = Y(A_i)$  for areas  $A_i, i = 1, \dots, n$ . The S-ARCH model is defined by

$$\begin{aligned} y_i &= \sigma_i \varepsilon_i \\ \log \sigma_i^2 &= \alpha + \rho \sum_{j=1}^n w_{i,j} \log y_j^2, \end{aligned}$$

where  $\varepsilon_i$  is an independent and identically distributed random variable with mean 0 and variance 1,  $w_{i,j} \geq 0$  is a spatial weight that quantifies a closeness from  $A_i$  to  $A_j$ ,  $\alpha$  and  $\rho$  is parameters and  $\rho$  is the parameter that describes the strength of spatial dependence of volatility on surrounding observations.

S-ARCH models are re-expressed in the form of spatial autoregressive (SAR) models for logged observations. Substituting  $\log \sigma_i^2$  into the log squared first equation, we have

$$\log y_i^2 = \alpha + \rho \sum_{j=1}^n w_{i,j} \log y_j^2 + \log \varepsilon_i^2,$$

namely SAR models.

The two parameters are estimated separately by a two step procedure in order to avoid the bias caused by joint estimation by least squares. In the first stage, we estimate only the parameter  $\rho$  by the least square method. Since the error term  $\log \varepsilon_i^2$  in the model is not a zero-mean process, least square estimator for  $\alpha$  would be biased. In the second stage, we consider the estimation of  $\alpha$ . Regarding  $\varepsilon$  in the error term as independent standard Gaussian variable, we estimate it by maximizing the quasi log-likelihood. Both the estimators for  $\rho$  and  $\alpha$  are consistent, while only the estimator for  $\rho$  can be proved to be asymptotically normal.

We carry out simulation studies to evaluate the finite sample properties of the estimators. We consider the S-ARCH model with the following four cases of i.i.d error term  $\varepsilon_i$  that follows

- normal distribution,
- Student-t distribution with 3 degrees of freedom,
- chi-squared distribution with 1 degrees of freedom,
- log-normal distribution.

Notice that all the error distributions were normalized to be mean 0 and variance 1. We find that the estimator for  $\rho$  has almost the same means and root of mean squared errors for each of the four error distributions, while the one for  $\alpha$  has empirical means and root of mean squared errors dependent on the error

distributions. As the error distribution is more discrepant from Gaussianity, the empirical estimation performance for  $\alpha$  is less efficient.

Finally, we apply S-ARCH models to land price data in Tokyo area. Land price data were collected by the Japanese Ministry of Land, Infrastructure, Transport, and Tourism, which publishes land prices (yen /  $m^2$ ) on many points scattered all over Japan regularly every year. Dividing Tokyo area into 347 discrete units consisting of wards, cities and towns, we evaluated the simple average of logged returns included in each unit, resulting in areal data of 347 observations from 2003 till 2014. We find first the significant evidence of S-ARCH effects in each year. Precisely, in the testing problem of  $H_0 : \rho = 0, H_1 : \rho \neq 0$ , the t values for  $\rho$  are all 5 % significant except for that in 2003. Moreover, we find the so called volatility clustering, namely units with higher spatial volatilities are clustered in some specific districts. The identified volatilities in 2010, 2011 and 2012, indicates that spatial volatilities in the coastal areas hit by the Tsunami by the Great East Japan Earthquake in 2011 are higher than those in the other two years. The identified volatilities in 2005-2010, indicates that spatial volatilities in central Tokyo grow in 2005-2007, the period of the economics boom, while they are almost extinct in 2008-2010, the period of the recession after Lehman shock. These behaviors of spatial volatilities suggest that spatial volatilities react to economic booms or recessions in the opposite way with time series volatilities, in recalling that typical financial time series volatilities burst in an economic shock while relatively stable in a boom.

In chapter 3, we extend a generalized autoregressive conditional heteroscedasticity (GARCH) model for time series to that for spatial data, which we call a spatial GARCH (S-GARCH) model. Suppose we observe spatial data  $y_i$  on a spatial area  $i$  for  $i = 1, \dots, n$ . We shall define S-GARCH models to describe spatial volatilities of  $y_i$  by

$$\begin{aligned} y_i &= \sqrt{h_i} \varepsilon_i, \\ \log h_i &= \lambda \sum_{j=1}^n w_{i,j} \log h_j + \rho \sum_{j=1}^n w_{ij} \log y_j^2 + \alpha + \mathbf{z}_i' \delta, \end{aligned}$$

where  $\sqrt{h_i}$  is volatility,  $\varepsilon_i$  is an independent and identically distributed (i.i.d) random variable with mean zero and variance 1,  $\mathbf{z}_i$  is  $(k \times 1)$  non-stochastic regressors, and  $w_{ij}$  is a spatial weight that quantifies vicinity from area  $i$  to area  $j$ . The matrix  $W_n$ ,  $n$  by  $n$  matrix composed of  $w_{ij}$ , is called a spatial weight matrix. A spatial weight matrix is usually determined by geographical information of spatial data. The first order contiguity weight matrix is a standard choice for it (Sato and Matsuda, 2017). For parameters  $(\lambda, \rho, \alpha, \delta)'$  in this model,  $\lambda$  and  $\rho$  describes spatial interactions of volatilities and logged observations. S-GARCH models reduce to S-ARCH models proposed by Sato and Matsuda (2017) when  $\lambda$  is equal to 0.

Let us re-express S-GARCH models as SARMA models. Denoting  $\log \mathbf{y}^2 = (\log y_1^2, \dots, \log y_n^2)'$ ,  $\log \mathbf{h} = (\log h_1, \dots, \log h_n)'$ ,  $\log \boldsymbol{\varepsilon}^2 = (\log \varepsilon_1^2, \dots, \log \varepsilon_n^2)'$ ,  $\mathbf{Z}_n = (\mathbf{z}_1, \dots, \mathbf{z}_n)'$ ,  $\mathbf{1}_n = (1, \dots, 1)'$  and  $I_n$  is  $n \times n$  identity matrix, we have the vector representation for S-GARCH models. Substituting second equation into first



equation, we have

$$\log \mathbf{y}^2 = (\lambda + \rho)W_n \log \mathbf{y}^2 + \alpha \mathbf{1}_n + Z_n \delta + (I_n - \lambda W_n) \log \boldsymbol{\varepsilon}^2,$$

which is a SARMA model, one of popular spatial econometrics models.

Parameters are estimated by a two step procedure. First step is the estimation of  $(\lambda, \rho, \delta)'$  by QML method. The constant term  $\alpha$  shall be estimated separately in the second step, as  $\log \boldsymbol{\varepsilon}^2$  is not zero mean and the estimator for constant term in the first step is biased. In second step,  $\alpha$  is estimated by the likelihood different from the one in the first step. Both estimators have consistency and asymptotic normality.

Monte Carlo experiments are carried out to investigate finite sample performances of the two stage estimators. we simulate areal data by S-GARCH models, where  $z_i$ 's are randomly generated from independent normal distributions and the spatial weights matrix is generated according to Rook contiguity and row normalizing. For the error terms,  $\varepsilon_i$ , we consider the three cases: (i) standard normal distributions, (ii) chi-squared distributions with 3 degrees of freedom and (iii) log normal distributions. The results show the estimators in the first step,  $(\hat{\lambda}, \hat{\rho}, \hat{\beta})'$  are nearly unbiased and not sensitive to the choice of the error distributions. On the other hand, the second step estimator,  $\hat{\alpha}$  depends on the error distribution, ie, is more biased and has larger RMSE for more deviations from Gaussian.

We shall apply S-GARCH models to land price data in Tokyo area in order to demonstrate identification of spatial volatilities. We use prefectural land price research as land price data. The Japanese Ministry of Land, Infrastructure, Transport, and Tourism publishes land prices on sampling points scattered irregularly all over Japan in the form of price per  $m^2$  in July. We focus on the land prices over Tokyo area (Tokyo, Kanagawa, Saitama, Chiba, Tochigi, Ibaraki, Gunma) observed yearly from 2009 to 2014. Averaging log returns of land price in municipal units, we obtain land price as yearly observations of areal data. Namely, land price data consists of yearly averaged log returns over 344 discrete areal unit's from 2010 to 2014. we find that  $\hat{\lambda}$ , the strength of interactions among spatial volatilities, are significant after the Great East Japan Earthquake in 2011 until 2013. This suggests that spatial volatility in land prices may have strengthened when the big event occurs.  $\hat{\rho}$ , the strength between spatial volatility and logged squared observations, is as large as the estimator for that of time series GARCH models. It is seen that  $\hat{\lambda} + \hat{\rho}$  is estimated to be close to 1 between 2011 and 2013, which likely causes volatility clustering around coastal areas. We find that volatilities not only at coastal areas hit by the Tsunami but also areas near Fukushima are identified to be high and , which suggests the effects of Fukushima nuclear accidents. Finally in comparison between identified volatilities by fitting S-ARCH and S-GARCH models, we observe that S-GARCH fits better in terms of AIC with global spillover, which means the identified volatility by S-GARCH models are more highly spatially correlated than those of S-ARCH models.

In chapter 4, we propose spatio-temporal extensions of time series multivariate volatility models. We call spatiotemporally extended volatility models

as spatial autoregressive models with spatial autoregressive error and generalized autoregressive conditional heteroskedasticity processes, namely SARAR-GARCH models. Let  $r_{i,t}$  be returns of financial instruments. We shall define SARAR-GARCH models to describe volatilities of return series  $r_{i,t}$  by

$$\begin{aligned} \mathbf{r}_t &= \lambda W \mathbf{r}_t + \mathbf{u}_t \\ \mathbf{u}_t &= \rho W \mathbf{u}_t + \boldsymbol{\varepsilon}_t \\ \varepsilon_{i,t} &= \sigma_{i,t} f_{i,t}, \\ f_{i,t} &\sim i.i.d(0, 1), \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i \sigma_{i,t-1}^2 + \beta_i \varepsilon_{i,t-1}^2, \end{aligned}$$

where  $\mathbf{r}_t = (r_{1,t}, \dots, r_{n,t})$ ,  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})$ ,  $\sigma_{i,t}$  is volatility,  $f_{i,t}$  is an independent and identically distributed (i.i.d) random variable with mean zeros and variance 1. The matrix  $W$ ,  $n$  by  $n$  matrix, is called a spatial weight matrix and pre-determined before analysis. For parameters  $(\rho, \lambda, \omega_i, \alpha_i, \beta_i)'$ ,  $\rho$  and  $\lambda$  describes spatial interactions of return series and  $\omega_i, \alpha_i$  and  $\beta_i$  are GARCH parameters. The positivity of  $\sigma_{i,t}^2$  is ensured by the following sufficient restrictions:  $\omega_i > 0$ ,  $\alpha_i \geq 0, \beta_i \geq 0$ , and the sum  $\alpha_i + \beta_i < 1$  for stationarity. It has been known that SARMA is guaranteed to exist when  $|\lambda| + |\rho| < 1$ .

A spatial weight matrix is usually determined by geographical information of spatial data and predetermined such as first-order contiguity relation or inverse distance between observations. However,  $r_{i,t}$  is financial data and doesn't include geographical information. Therefore, we need to determine financial distances to make a spatial weight matrix. Some author have proposed spatial weight matrix based on financial distance calculated from financial statement data such as dividend yields or market capitalizations. Here, we propose a method to make spatial weight matrices from financial data by stepwise backward regression. we apply the multiple linear regression model:

$$r_{i,t} = \delta_0 + \sum_{i \neq j}^n \delta_j r_{j,t} + z_{i,t},$$

where  $z_{i,t}$  follows i.i.d normal distribution. Then, we obtain the least square estimates  $\delta_j$  and those t-values. After that we check the minimum t-values and If the value is smaller than a critical value, for example 1.96, then we remove observations which have minimum t-value. Next, we regress  $r_{i,t}$  on  $n-2$  observations and we repeat this procedure until the minimum t-value is greater than the critical value.

We shall propose estimation of the parameters  $(\rho, \lambda, \beta_{i,j})'$  in SARAR-GARCH models. Parameters are estimated by a two step procedure. First step is the estimation of  $\lambda$  and  $\rho$  and second step is that of  $\beta_{i,j}$ . Parameter  $\rho$  and  $\lambda$  are estimated in first step. We regard  $\sigma_{i,t}$  as constant variance and we apply quasi-maximum likelihood method with the model. After that we apply GARCH models with residuals derived from first step in second step.

In real data analysis, We apply the SARAR-GARCH model to daily returns of the Nikkei 225 stock price data and S&P 500 stock price data, that is the returns  $(r_{i,t})$  are computed as  $100(\log P_t - \log P_{t-1})$ , where  $P_t$  is the closing price and  $t$  is the time index referring to trading day  $t$ . We compare the in-sample and out-sample performances of SARAR-GARCH models with those of CCC models which is a benchmark. First, we check the in-sample performances based on log-likelihood. The results show the log-likelihood of the CCC model is grater than that of SARAR-GARCH. This means model fitting of the CCC model is better. One reason is that the number of parameters in CCC models is more than five times of those of SARAR-GARCH models. Secondly, we compare out-sample performances. The results shows out-sample performance of SARAR-GARCH models are better. This shows CCC model may be over-fitting and it cause lower forecasting performance. Moreover, we find SARAR-GARCH models work quite well in U.S market analysis. The reason why proposed models work well in U.S market is stock prices in U.S market are more volatile. CCC models assume constant correlation between stock prices so can't capture dynamic relations, but SARAR-GARCH models can capture dynamic correlation as volatility matrix for the model shown . Therefore, SARAR-GARCH models work well in U.S market.

# Bibliography

- [1] Alexander, C. (2001). Market models: A guide to financial data analysis. John Wiley & Sons.
- [2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
- [3] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics*, 498-505.
- [4] Borovkova, S. & Lopuhaa, R., (2012). Spatial GARCH: A Spatial Approach to Multivariate Volatility Modeling. Available at SSRN: <https://ssrn.com/abstract=2176781>.
- [5] Caporin, M., & Paruolo, P., 2009, Structured multivariate volatility models, Available at SSRN: <http://ssrn.com/abstract=1318639>.
- [6] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.
- [7] Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1), 1-50.
- [8] Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory*, 11(1), 122-150.
- [9] Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5), 1779-1801.
- [10] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.
- [11] Robinson, P. M. (2009). Large- sample inference on spatial dependence. *The Econometrics Journal*, 12(s1).
- [12] Sato, T. & Matsuda, Y. (2017). Spatial Autoregressive Conditional Heteroskedasticity Models. *J.Japan Statist. Soc.*, Vol. 47, 2

- [13] Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, 20(3), 351-362.
- [14] Yan, J. (2007). Spatial stochastic volatility for lattice data. *Journal of agricultural, biological, and environmental statistics*, 12(1), 25.

## Chapter 2

# SPATIAL AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY MODELS

### Abstract

This study proposes a spatial extension of time series autoregressive conditional heteroskedasticity (ARCH) models to those for areal data. We call the spatially extended ARCH models as spatial ARCH (S-ARCH) models. S-ARCH models specify conditional variances given surrounding observations, which constitutes a good contrast with time series ARCH models that specify conditional variances given past observations. We estimate the parameters of S-ARCH models by a two-step procedure of least squares and the quasi maximum likelihood estimation, which are validated to be consistent and asymptotically normal. We demonstrate the empirical properties by simulation studies and real data analysis of land price data in Tokyo areas.

### 2.1 Introduction

This paper aims to extend autoregressive conditional heteroskedasticity (ARCH) models for time series by Engle (1982) and Bollerslev (1986) to those for areal data, which we call spatial ARCH (S-ARCH) models. Areal data is a kind of spatial data that is composed of discrete observations on areal units. See Figure 2.1 for an example of areal data of land prices in Tokyo areas, which are

yearly returns in 2014 on areal units of wards, cities and towns. A S-ARCH model is the one that would describe a conditional variance at an areal unit given data at all other areal units, which we shall call spatial volatility, while a time series ARCH model gives a model to express a conditional variance given past observations, which we call time series volatility. Spatial and time series volatility will be distinguished strictly in this paper. Robinson (2009) has done a spatial extension of stochastic volatility (SV) models for time series (Taylor (2008)). This paper can be regarded as an alternative trial of spatial extensions of time series volatility models in terms of ARCH models.

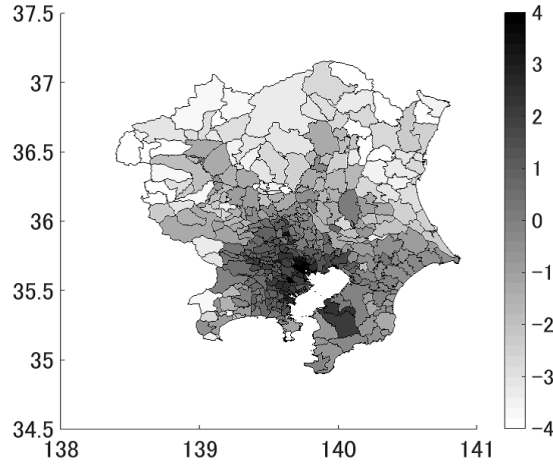


Figure 2.1: Yearly log returns of the land prices over 347 areal units of wards, cities and towns in Tokyo areas (Tokyo, Kanagawa, Saitama, Chiba, Tochigi, Ibaraki and Gunma prefectures) in 2014.

It is often observed that spatial volatility is not a constant but depends on data at surrounding areal units similarly like time series volatility in financial time series. It is well known that financial time series, even after whitened such as by autoregressive (AR) models, often exhibit substantial dependency in the sense that squared series is serially correlated, which has been accounted for by time series ARCH models (see Tsay (2005)). Hence serial correlations of squared residuals are checked to detect heteroskedasticity of time series volatility.

Let us demonstrate heteroskedasticity of spatial volatility for areal data in the same way as in financial time series case. Land price data in Tokyo areas observed yearly by the Japanese Ministry of Land, Infrastructure, Transport, and Tourism is employed for the demonstration. Fitting spatial autoregressive (SAR) models (see such as Ord (1975), Kelejian (1999), Lee (2004) and LeSage

and Pace (2009)) to yearly log returns of land price data from 2009 till 2011 to remove linear dependency, we obtain the residuals from the original areal data for each year. Table 2.1 shows Moran's I's of the squared residuals as well as those of the residuals, where Moran's I is an index of spatial correlation that can be regarded as a spatial analogue of lag 1 autocorrelation in time series (Moran (1948, 1950), Anselin (1988)). We find that substantial dependency for the squared residuals still remains, while linear dependency of the residuals is almost null, which suggests heteroskedasticity of spatial volatility exists and that S-ARCH models may work to account for it.

The interesting features of S-ARCH models are summarized as follows. First, the existence condition of S-ARCH models is easily established. Usually it is difficult to check the condition, since S-ARCH models are not Markovian but Markov random fields and so the techniques for Markov models including ARCH models cannot be applied. Secondly, estimators for the parameters in S-ARCH models are well validated to be consistent and asymptotic normally distributed.

The paper proceeds as follows. Section 2.2 introduces S-ARCH model. Section 2.3 shows the estimation procedures and their asymptotic properties. Section 2.4 examines empirical properties of S-ARCH models by applying them to simulated and land price data in Tokyo area. Section 6 discusses some concluding remarks.

## 2.2 Spatial Autoregressive Conditional Heteroskedasticity model

Suppose we have areal data  $y_i = y(A_i)$  for areas  $A_i, i = 1, \dots, n$ . Let us introduce a spatial autoregressive conditional heteroskedasticity (S-ARCH) model for areal data  $y_i$ . The S-ARCH model is defined by

$$y_i = \sigma_i \varepsilon_i \quad (2.1)$$

where  $\varepsilon_i$ s are independent and identically distributed random variables with mean 0 and variance 1, and  $\sigma_i$  satisfies the following relation:

$$\log \sigma_i^2 = \alpha + \rho \sum_{j=1}^n w_{ij} \log y_j^2, \quad (2.2)$$

where  $w_{ij} \geq 0$  is a spatial weight that quantifies a closeness from  $A_i$  to  $A_j$  with  $w_{ii} = 0$ , and constitutes a spatial weight matrix  $W = (w_{ij})$ . One popular choice

Table 2.1: Moran's I's for the residuals and the squared residuals after fitting SAR models to log returns of land price data in Tokyo areas from 2009 till 2011.

	2009	2010	2011	2012
original	-0.043	-0.010	-0.036	-0.075
squared	0.152	0.322	0.295	0.246



of the weights  $w_{ij}$  is the one that takes 1 when  $A_i$  and  $A_j$  share the common border and 0 otherwise.  $W$  may not necessarily be dependent on physical distance among areas but on any other abstract closeness such as similarity of financial conditions and so on (Beck et al. (2006)).  $\rho$  is the parameter that describes the strength of spatial dependence of volatility on surrounding observations.

S-ARCH models are different from time series ARCH models in the following two points. Firstly, spatial volatility at one areal unit is described by observations at all other units, which is different from time series volatility descriptions in ARCH models following time flows from past to future. Figure 2.2 is a simple example of dependence structures of spatial data demonstrating that spatial volatility at one unit in S-ARCH models can be influenced from observations at all other units. Although time series and spatial volatilities are defined in the different ways, we have found in this paper that they share the similar properties such as significant correlations of squared or absolute processes and the so called volatility clustering which means one large values tend to induce large surrounding and future values in spatial and time series cases, respectively.

Next, the log transformation of  $\sigma_i$  is used to ensure the existence of areal data  $y_i$ . If we defined the spatial volatility similarly with time series ARCH models by

$$\sigma_i^2 = \alpha + \rho \sum_{j=1}^n w_{ij} y_j^2,$$

it would be very difficult to guarantee existence of areal data  $y_i$  unlike that for time series ARCH models that can be proved by Markov process theories (Fan and Yao (2003)). The log transformation of  $\sigma_i^2$  makes it much easier to prove the existence in the following way. Substituting  $\log \sigma_i^2$  in the log squared equation of (2.1) with (2.2), we have

$$\log y_i^2 = \alpha + \rho \sum_{j=1}^n w_{ij} \log y_j^2 + \log \varepsilon_i^2, \quad (2.3)$$

which is a spatial autoregressive (SAR) model whose existence conditions have been well established. When the matrix  $I - \rho W$  is non-singular,  $y_i$  is guaranteed to exist as

$$\log \mathbf{y}^2 = (I - \rho W)^{-1} \alpha \mathbf{1} + (I - \rho W)^{-1} \log \boldsymbol{\varepsilon}^2,$$

where  $\log \mathbf{y}^2 = (\log y_1^2, \dots, \log y_n^2)'$ ,  $\mathbf{1} = (1, \dots, 1)'$  and  $\log \boldsymbol{\varepsilon}^2 = (\log \varepsilon_1^2, \dots, \log \varepsilon_n^2)'$ . When  $W$  is row normalized, which means the sum of each row of the matrix is normalized to be 1,  $I - \rho W$  is non-singular if  $|\rho| < 1$ . When  $W$  is not row normalized, the inverse exists if  $\rho \in (\frac{1}{\lambda_{(n)}}, \frac{1}{\lambda_{(1)}})$  where  $\lambda_{(1)} < \dots < \lambda_{(n)}$  are the ordered eigenvalues of  $W$  (Banerjee et al. (2014), Arbia (2014)).

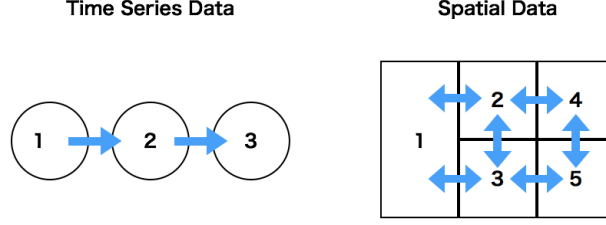


Figure 2.2: Simple example of spatial dependency for 5 areal units. Spatial volatility at one unit can be influenced by observations at all neighboring units simultaneously. Compare time series case when only past observations can influence time series volatility in ARCH models.

## 2.3 Estimation

We consider the estimation of the two parameters  $\alpha, \rho$  in (2.2) and the asymptotic properties of the estimators. The two parameters will be estimated separately by a two step procedure in order to avoid the bias caused by joint estimation by least squares. We shall show that both the estimators for  $\rho$  and  $\alpha$  are consistent, while only the estimator for  $\rho$  can be proved to be asymptotically normal.

### 2.3.1 Parameter estimation by the two step procedure

The parameters  $\alpha, \rho$  in (2.2) will be estimated separately in the two step procedure. Since the error term  $\log \varepsilon_i^2$  in the model (2.3) is not a zero-mean process, joint estimation of  $\alpha$  and  $\rho$  by the usual least squares (LS) would not work. Specifically the LS estimator for  $\alpha$  would be biased because of the non-zero mean error terms. In the first stage we estimate only the parameter  $\rho$  by the LS method, while in the second stage we estimate  $\alpha$  by the quasi maximum likelihood method by regarding  $\varepsilon_i$  in the error term as a standard normal variable.

Let us begin from the LS estimation of  $\rho$  in the first stage. We slightly modify the original form in (2.3) as

$$\log y_i^2 = \alpha + E \log \varepsilon_i^2 + \rho \sum_{j=1}^n w_{ij} \log y_j^2 + \log \varepsilon_i^2 - E \log \varepsilon_i^2.$$

Define three symbols for simplicity by

$$\begin{aligned} z_i &= \log y_i^2, \\ \phi &= \alpha + E \log \varepsilon_i^2, \\ v_i &= \log \varepsilon_i^2 - E \log \varepsilon_i^2. \end{aligned}$$

Then we have

$$z_i = \rho \sum_{j=1}^n w_{ij} z_j + \phi + v_i. \quad (2.4)$$

Let us estimate  $\rho$  in (2.4) by the least squares, which will be obtained by regarding  $v_i$ s as independent Gaussian variables with mean 0. Denoting the variance of  $v_i$  as  $\sigma^2$ , we have the quasi log-likelihood function of  $\theta = (\phi, \rho, \sigma^2)'$  by

$$\begin{aligned} \log L_n(\theta) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) + \log \|I_n - \rho W\| \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^n \left( z_i - \rho \sum_{j=1}^n w_{ij} z_j - \phi \right)^2, \end{aligned} \quad (2.5)$$

where  $\|I_n - \rho W\|$  is the Jacobian term.

Concentrating out  $\phi, \sigma^2$  by

$$\begin{aligned} \hat{\phi}_n(\rho) &= \frac{1}{n} \sum_{i=1}^n \left( z_i - \rho \sum_{j=1}^n w_{ij} z_j \right), \\ \hat{\sigma}_n^2(\rho) &= \frac{1}{n} \sum_{i=1}^n \left( z_i - \hat{\phi}_n(\rho) - \rho \sum_{j=1}^n w_{ij} z_j \right)^2. \end{aligned}$$

we have the concentrated log-likelihood given by

$$\log L_n(\rho) = -\frac{n}{2} (\log(2\pi) + 1) - \frac{n}{2} \log \hat{\sigma}_n^2(\rho) + \log \|I_n - \rho W\|. \quad (2.6)$$

Maximizing this with respect to  $\rho$ , we have the estimator  $\hat{\rho}_n$ .

Next in the second stage, let us consider the estimation of  $\alpha$ . Regarding  $\varepsilon_i$ s in the error term in (2.3) as independent standard Gaussian variables, we estimate it by maximizing the quasi log-likelihood. If  $\epsilon_i$  follows a standard normal distribution, a probability density function of  $\log \varepsilon_i^2$  is easily obtained via change of variables formula as

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \exp(x) + \frac{1}{2} x\right). \quad (2.7)$$

Figure 2.3 shows the probability density function (pdf) of  $\log \epsilon_i^2$ .

Then the quasi log-likelihood function of (2.3) based on the density (2.7) is given by

$$\begin{aligned} \log L(\alpha, \rho) &= \sum_{i=1}^n \left\{ \frac{y_i^2}{\exp(\alpha + \rho \sum_{j=1}^n w_{ij} z_j)} - \alpha + \rho \sum_{j=1}^n w_{ij} z_j \right\} \\ &\quad - 2 \log \|I_n - \rho W\|. \end{aligned} \quad (2.8)$$

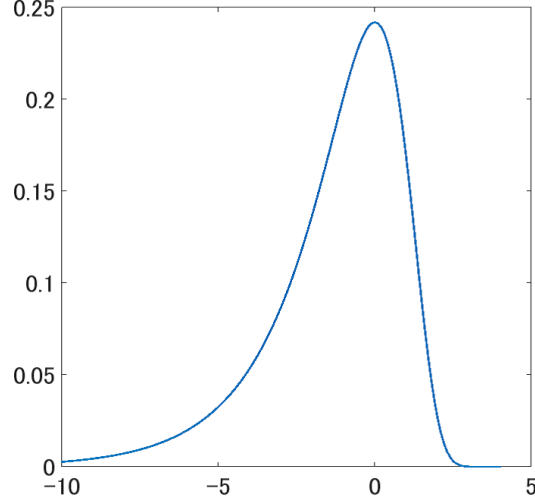


Figure 2.3: The probability density function of  $\log \varepsilon_i^2$  when  $\varepsilon_i$  follows a standard normal distribution.

Differentiating it with respect to  $\alpha$ , we have the quasi maximum likelihood estimator given by

$$\alpha = \log \left\{ \frac{1}{n} \sum_{i=1}^n \exp \left( z_i - \rho \sum_{j=1}^n w_{ij} z_j \right) \right\}. \quad (2.9)$$

Substituting  $\rho$  with our proposed estimator  $\hat{\rho}_n$  in the first stage, we propose

$$\hat{\alpha}_n = \log \left\{ \frac{1}{n} \sum_{i=1}^n \exp \left( z_i - \hat{\rho}_n \sum_{j=1}^n w_{ij} z_j \right) \right\}, \quad (2.10)$$

as an estimator for  $\alpha$ .

It is possible to estimate  $\rho$  and  $\alpha$  jointly by maximizing the quasi likelihood in (2.8). The quasi MLEs for them are, however, not validated to be consistent. That is the reason why we propose the two step procedure for the estimation that will be established to be consistent for  $\hat{\rho}_n$  and  $\hat{\alpha}_n$  and asymptotically normal for  $\hat{\rho}_n$  in the next section.

### 2.3.2 Asymptotic results

This section considers the consistency of  $\hat{\rho}_n$  maximizing (2.6) and  $\hat{\alpha}_n$  in (2.10), and later derives asymptotic normality of  $\hat{\rho}_n$ . Consistency and asymptotic normality of  $\hat{\rho}_n$  are proved by the results of Lee (2004). The consistency of  $\hat{\alpha}_n$  is

easily proved independently of  $\hat{\rho}_n$ , while we have not succeeded in deriving the asymptotic normality.

We will make use of the following set of assumptions. Let  $\rho_0$  and  $\alpha_0$  be the true values of  $\rho$  and  $\alpha$ , respectively.

**Assumption 1.**  $\{\log \varepsilon_i^2\}$ ,  $i=1, \dots, n$ , are i.i.d with finite mean and variance. Its moment  $E(|\log \varepsilon^2|^{4+\gamma})$  for some  $\gamma \geq 0$  exists.

**Assumption 2.** The  $W_n$  is a row-normalized matrix and uniformly bounded in column sums, in other words for some real constant  $c$ ,  $\sum_{i=1}^n w_{ij} < c$  for all  $j$ .

**Assumption 3.** The matrix  $(I_n - \rho_0 W)$  is nonsingular and  $(I_n - \rho_0 W)^{-1}$  is uniformly bounded in both row and column sums.

**Assumption 4.**  $(I_n - \rho W)^{-1}$  is uniformly bounded in either row or column sums, uniformly in  $\rho$  in a compact parameter space  $\Theta = [-\delta, \delta]$ , where  $\delta$  is smaller than 1. The true  $\rho_0$  is in the interior of  $\Theta$ .

**Assumption 5.** The  $\lim_{n \rightarrow \infty} (G_n \mathbf{1}_n \beta_0)' \mathbf{1}_n / \mathbf{n}$  and  $\lim_{n \rightarrow \infty} (G_n \mathbf{1}_n \beta_0)' (\mathbf{G}_n \mathbf{1}_n \beta_0) / \mathbf{n}$  exist and are nonsingular, where  $G_n = W_n (I_n - \rho_0 W_n)^{-1}$  and  $\mathbf{1}_n$  is a vector whose elements are all 1.

First, we consider consistency and asymptotic normality of  $\hat{\rho}_n$ .

The covariance matrix of  $(1/n) \partial \log L_n(\theta_0) / \partial \theta$  is

$$E \left( \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\theta_0)}{\partial \theta'} \right) = -E \left( \frac{\partial^2 \log L_n(\theta_0)}{\partial \theta \partial \theta'} \right) + \Omega_{\theta, n},$$

where the average Hessian matrix is

$$-E \left( \frac{1}{n} \frac{\partial^2 \log L_n(\theta_0)}{\partial \theta \partial \theta'} \right) = \begin{pmatrix} \frac{1}{\sigma_0^2} & \frac{1}{\sigma_0^2 n} \mathbf{1}_n' (G_n \mathbf{1}_n \phi) & 0 \\ * & \frac{1}{\sigma_0^2 n} (G_n \mathbf{1}_n \phi)' (G_n \mathbf{1}_n \phi) + \frac{1}{n} \text{tr}((G_n G_n') G_n) & \frac{1}{\sigma_0^2 n} \text{tr}(G_n) \\ * & * & \frac{1}{2\sigma_0^4} \end{pmatrix},$$

and

$$\Omega_{\theta, n} = \begin{pmatrix} 0 & \frac{\mu_3}{\sigma_0^4 n} \sum_{i=1}^n G_{n, ii} & \\ * & \frac{2\mu_3 \phi}{\sigma_0^4} \sum_{i=1}^n G_{n, ii} G_{i, n} + \frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4 n} \sum_{i=1}^n G_{n, ii}^2 & \\ * & * & \frac{\mu_3}{2\sigma_0^6} \\ & \frac{1}{2\sigma_0^6 n} [\phi \mu_3 \sum_{i=1}^n \sum_{j=1}^n G_{n, ij} + (\mu_4 - 3\sigma_0^4) \text{tr}(G_n)] & \\ & \frac{(\mu_4 - 3\sigma_0^4)}{4\sigma_0^8} & \end{pmatrix},$$

is a symmetric matrix with  $\mu_j = E(v_i^j)$ ,  $j = 3, 4$ , being the third, and fourth moments of  $v_i$ , where  $G_{in}$  is the  $i$ th row of  $G_n$ ,  $G_{n, ij}$  is the  $(i, j)$ th entry of  $G_n$ .

**Theorem 1.** Under Assumptions 1-5,  $\hat{\rho}_n$  converges to  $\rho_0$  in probability and  $\sqrt{n}(\hat{\rho}_n - \rho_0) \xrightarrow{D} N(0, \sigma_\rho)$ ,  $\sigma_\rho$  is the  $(2, 2)$  element of matrix  $\Sigma_\theta^{-1} + \Sigma_\theta^{-1} \Omega_\theta \Sigma_\theta^{-1}$ , where  $\Omega_\theta = \lim_{n \rightarrow \infty} \Omega_{\theta, n}$  and  $\Sigma_\theta = -\lim_{n \rightarrow \infty} E \left( \frac{1}{n} \frac{\partial^2 \log L_n(\theta_0)}{\partial \theta \partial \theta'} \right)$ .

PROOF. We show Assumptions of S-ARCH models satisfy the 8 Assumptions of Lee (2004). Assumption 1 of S-ARCH models coincides with Assumption 1 of Lee (2004). Assumption 2 of S-ARCH models suffices Assumptions 2 and 3 of Lee (2004). The first half of Assumption 3 of S-ARCH models is the same as Assumption 4 of Lee (2004). Assumptions 2 and 3 of S-ARCH models suffice the assumption 5 of Lee (2004). The equation (2.4) has an only intercept term as explanatory variables. Therefor, Assumption 6 of Lee (2004) holds. Assumption 4 of S-ARCH models correspond with Assumption 7 of Lee (2004). Assumption 5 of S-ARCH models coincides with Assumption 8 of Lee (2004). Thus,  $\hat{\rho}_n$  is a consistent and asymptotically normal.

Secondly, we consider consistency of  $\hat{\alpha}_n$

**Theorem 2.** *Under Assumptions 1-5,  $\hat{\alpha}_n$  converges to  $\alpha_0$  in probability.*

PROOF. The consistency of  $\hat{\alpha}_n$  will follow from the convergence in probability to zero of  $(\exp(\hat{\alpha}_n) - \exp(\alpha_0))$ .

From (2.10), we have

$$\exp(\hat{\alpha}_n) = \frac{1}{n} \sum_{i=1}^n \exp\left(z_i - \hat{\rho} \sum_{j=1}^n w_{ij} z_j\right).$$

Let  $\log \epsilon_i^2$  be  $\zeta_i$  for simplicity. Thus,

$$\begin{aligned} \exp\left(z_i - \hat{\rho} \sum_{j=1}^n w_{ij} z_j\right) &= \exp\left\{z_i - \rho_0 \sum_{j=1}^n w_{ij} z_j + (\rho_0 - \hat{\rho}) \sum_{j=1}^n w_{ij} z_j\right\}, \\ &= \exp\left(z_i - \rho_0 \sum_{j=1}^n w_{ij} z_j\right) \left\{ \exp\left((\rho_0 - \hat{\rho}) \sum_{j=1}^n w_{ij} z_j\right) - 1 + 1 \right\}, \\ &= \exp(\alpha_0 + \zeta_i) + \\ &\quad \exp(\alpha_0 + \zeta_i) \left\{ \exp\left((\rho_0 - \hat{\rho}) \sum_{j=1}^n w_{ij} z_j\right) - 1 \right\}. \end{aligned}$$

By Theorem 1,  $(\rho_0 - \hat{\rho}) \sum_{j=1}^n w_{ij} z_j = (\rho_0 - \hat{\rho})(z_i - \alpha_0 - \zeta_i) = o_p(1)O_p(1) = o_p(1)$ . Since  $\exp((\rho_0 - \hat{\rho}) \sum_{j=1}^n w_{ij} z_j) \xrightarrow{p} 1$ ,  $\exp((\rho_0 - \hat{\rho}) \sum_{j=1}^n w_{ij} z_j) - 1 \xrightarrow{p} 0$ . Moreover,  $\exp(\alpha_0 + \zeta_i) = O_p(1)$ . Therefore,

$$\begin{aligned} \exp\left(z_i - \hat{\rho} \sum_{j=1}^n w_{ij} z_j\right) &= \exp(\alpha_0 + \zeta_i) + O_p(1)o_p(1), \\ &= \exp(\alpha_0 + \zeta_i) + o_p(1). \end{aligned}$$

Then,

$$\begin{aligned} \exp(\hat{\alpha}_n) &= \frac{1}{n} \sum_{i=1}^n \{\exp(\alpha_0 + \zeta_i) + o_p(1)\}, \\ &= \exp(\alpha_0) \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 + o_p(1). \end{aligned}$$

By the law of large numbers (Brockwell and Davis (2013)),

$$\exp(\hat{\alpha}_n) \xrightarrow{p} \exp(\alpha_0) + o_p(1).$$

Therefore,  $\exp(\hat{\alpha}_n) - \exp(\alpha_0) \xrightarrow{p} o_p(1)$ .

## 2.4 Empirical Analysis

We shall examine the empirical properties of S-ARCH models by applying to simulated and land price data in Tokyo areas. In the simulations we examine finite sample performances of the estimators, while in the real data analysis we check S-ARCH effects in land prices by testing if  $\rho$  is positive or zero and evaluate the spatial volatilities identified by S-ARCH models and show them graphically.

### 2.4.1 Simulation studies

Let us consider the finite sample properties of the estimators  $\hat{\rho}_n$  and  $\hat{\alpha}_n$  by simulated data. The disturbance term in the S-ARCH model is designed as several cases of non-Gaussian as well as Gaussian distributions to see the effects of discrepancy from Gaussianity.

We consider the S-ARCH model in (2.1) and (2.2) with the following four cases of iid error term  $\varepsilon_i$  that follows

- normal distribution (norm),
- Student-t distribution with 3 degrees of freedom (t(3)),
- chi-squared distribution with 1 degrees of freedom(chi(1)),
- log-normal distribution(log-norm).

Notice that all the error distributions were normalized to be mean 0 and variance 1. The spatial volatility was designed by (2.2) with  $\alpha = 0.3$  and  $\rho = 0.2$ , where  $W = (w_{ij})$ , the spatial weight matrix, was the row normalized first-order contiguity relation for 347 areal units in Tokyo areas whose map is in Figure 1, which will be employed again in the following section of land price data analysis. We have conducted estimation of  $\rho$  and  $\alpha$  by the two step procedure for 1000 sets of data simulated with each of the four cases of error terms to check the empirical performances of the estimators.

The empirical means and square root of mean squared errors (RMSE) for the two estimators are reported in Table 2.2. We find that the estimator for  $\rho$  has almost the same means and RMSEs for each of the four error distributions, while the one for  $\alpha$  has empirical means and RMSEs dependent on the error distributions. As the error distribution is more discrepant from Gaussianity, the empirical estimation performance for  $\alpha$  is less efficient. It follows that we have

to be careful for a negative bias in the estimation of  $\alpha$  when the error term is discrepant from Gaussianity that can cause the identified spatial volatility to be somewhat smaller than the actual ones.

Table 2.2: The empirical means and square root of mean squared errors (RMSE) of the estimators for  $\rho = 0.2$  and  $\alpha = 0.3$  evaluated by 1000 simulations.

distribution	mean( $\rho$ )	RMSE( $\rho$ )	mean( $\alpha$ )	RMSE( $\alpha$ )
norm	0.188	0.080	0.279	0.125
t(3)	0.189	0.079	0.206	0.329
chi(1)	0.183	0.083	0.245	0.237
log-norm	0.187	0.081	0.190	0.418

### 2.4.2 Land price data analysis

We apply S-ARCH models in (2.1) to land price data in Tokyo area. We shall examine whether the two properties that are typical in financial time series are detected in land price data. One is S-ARCH effect, which we mean dependencies of spatial volatility on surrounding observations, and the other is volatility clustering.

Land price data were collected by the Japanese Ministry of Land, Infrastructure, Transport, and Tourism, which publishes land prices (yen / $m^2$ ) on many points scattered all over Japan regularly every year. (<http://nlftp.mlit.go.jp/ksj/>). Two kinds of land price data are available in the website as prefectural land price research and public land price. Here we chose prefectural land price research published in July every year as data sources, and collected land prices on several thousand points scattered over Tokyo area (Tokyo, Kanagawa, Saitama, Chiba, Tochigi, Ibaraki, Gunma) observed from 2002 till 2014 and transformed them into logged returns from 2003 till 2014. Dividing Tokyo area into 347 discrete units consisting of wards, cities and towns, we evaluated the simple average of logged returns included in each unit, resulting in areal data of 347 observations from 2003 till 2014. Let us denote the logged returns at areal unit  $i$  for year  $t$  as  $y_{it}$ .

#### <<Table 2.3>>

In prior to S-ARCH model fitting to  $y_{it}$ , we apply year by year the following spatial autoregressive (SAR) model,

$$y_{it} = \beta + \kappa \sum_{j=1}^{347} w_{ij} y_{jt} + u_{it}, u_{it} \sim iid(0, \tau^2), \quad (2.11)$$

to remove spatial correlations. Here,  $W = (w_{ij})$ , the spatial weight matrix, is given by the row normalized first-order contiguity relation that takes 1 only when sharing common borders for 347 areal units in Figure 2.1. Table 2.3 shows



the estimated values of  $\hat{\beta}, \hat{\kappa}$  and  $\hat{\tau}^2$  in the SAR model in each year. We find the substantial correlations caused by  $\hat{\kappa}$  ranging around 0.8-0.9, which means strong similarities of land price returns among neighboring areal units.

<<Table 2.4>>

<<From Figure 2.4 to Figure 2.6>>

To the residual

$$\hat{u}_{it} = y_{it} - \hat{\beta} - \kappa \sum_{j=1}^{347} w_{ij} y_{jt},$$

which are obtained after fitting the SAR models, we applied the S-ARCH model in (2.3) year by year, where the same spatial weight matrix as that in the SAR model was employed. Table 2.4 shows the estimated values of  $\rho$  and  $\alpha$  in each year, where the standard errors of  $\hat{\rho}_n$ , which are derived in Theorem 2 by replacing the population moments with the corresponding sample ones, are also shown in parenthesis. Figures 2.4-2.6 express the spatial volatility evaluated by

$$\hat{\sigma}_i^2 = \exp \left( \hat{\alpha} + \hat{\rho} \sum_{j=1}^{347} w_{ij} \log(\hat{u}_j^2) \right).$$

We find first from Table 2.4 the significant evidence of S-ARCH effects in each year. Precisely, in the testing problem of  $H_0 : \rho = 0, H_1 : \rho \neq 0$ , the  $t$  values for  $\rho$  are all 5 % significant except for that in 2003. From Figures 2.4-2.5, we find the so called volatility clustering, namely units with higher spatial volatilities are clustered in some specific districts. Figure 2.4, which shows the identified volatilities in 2010, 2011 and 2012, indicates that spatial volatilities in the coastal areas hit by the Tsunami by the Great East Japan Earthquake in 2011 are higher than those in the other two years. Figures 2.5, which shows the identified volatilities in 2005-2010, indicates that spatial volatilities in central Tokyo grow in 2005-2007, the period of the economics boom, while they are almost extinct in 2008-2010, the period of the recession after Lehman shock. These behaviors of spatial volatilities suggest that spatial volatilities react to economic booms or recessions in the opposite way with time series volatilities, in recalling that typical financial time series volatilities burst in an economic shock while relatively stable in a boom.

## 2.5 Conclusion

We have proposed a spatial autoregressive conditional heteroskedasticity (S-ARCH) model to evaluate spatial volatility. Describing logged volatility with linear combinations of logged observations in S-ARCH models, we have established the conditions that guarantee the existence of S-ARCH models. Re-expressing S-ARCH models in the form of spatial autoregressive (SAR) models

for logged observations, we propose the two step procedure to estimate the parameters  $\rho$  and  $\alpha$  in S-ARCH models. Both the two estimators are validated to be consistent asymptotically, while the estimator for  $\rho$  is shown to be asymptotically normal as well as consistent. Finite sample performances of the procedure are reasonably good from our simulation studies. In the land price data analysis, we detect S-ARCH effects by testing if  $\rho$  is positive and find volatility clustering that reacts to economic shock oppositely with that in financial time series .

We complete the paper by describing possible extensions for future research. In the empirical analysis of land prices, we used the first-order contiguity relations as the spatial weight matrix. As (Beck et al. (2006)) shows, spatial distances that differ from geographic distances can be more interesting to improve our volatility analysis using S-ARCH models. We evaluated spatial volatility for fixed  $t$  only year by year. Spatio-temporal extensions of S-ARCH models would make it possible to analyze volatility jointly in space and time and to provide a more detailed analysis of volatility structures.

# Bibliography

- [1] Anselin, L. (1988). *Spatial Econometrics : Methods and Models*, Kluwer Academic Publishers, The Netherlands.
- [2] Arbia, G. (2014). *A Primer for Spatial Econometrics: with Applications in R*, Springer-Verlag, New York.
- [3] Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton.
- [4] Beck, N., Gleditsch, K. S., and Beardsley, K. (2006). Space is more than geography: Using spatial econometrics in the study of political economy, *International Studies Quarterly*, **50**(1), 27–44.
- [5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econ.*, **31**(3), 307–327.
- [6] Brockwell, P. J., and Davis, R. A. (2013). *Time Series: Theory and Methods*, Springer-Verlag, New York.
- [7] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.
- [8] Fan, J., and Yao, Q. (2003). *Nonlinear Time Series*, Springer-Verlag, New York.
- [9] Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model, *International economic review*, **40**(2), 509–533.
- [10] Lee, L. F. (2004). Asymptotic distributions of quasi- maximum likelihood estimators for spatial autoregressive models, *Econometrica*, **72**(6), 1899–1925.
- [11] LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*, CRC Press, Boca Raton.
- [12] Moran, P. A. (1948). The interpretation of statistical maps, *J. R. Stat. Soc. Ser. B.*, **10**(2), 243–251.

- [13] Moran, P. A. (1950). A test for the serial independence of residuals, *Biometrika*, **37(1-2)**, 178–181.
- [14] Ord, K. (1975). Estimation methods for models of spatial interaction. *J. Am. Statist. Assoc.*, **70(349)**, 120–126.
- [15] Robinson, P. M. (2009). Large- sample inference on spatial dependence, *Econom. J.*, **12(s1)**, S68-S82.
- [16] Taylor, S. J. (2008). *Modelling Financial Time Series*, world scientific, Singapore.
- [17] Tsay, R. S. (2005). *Analysis of Financial Time Series*, John Wiley & Sons, Hoboken.

Table 2.3: Estimated values and their standard errors for  $\kappa$ ,  $\beta$  and  $\tau^2$  in the spatial autoregressive (SAR) model in (2.11) applied to log returns of land price data in Tokyo year by year in 2003-2014. .

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
$\kappa$	0.865 (0.031)	0.836 (0.033)	0.805 (0.033)	0.830 (0.028)	0.933 (0.016)	0.848 (0.027)	0.880 (0.024)	0.700 (0.046)	0.820 (0.030)	0.831 (0.028)	0.880 (0.023)	0.892 (0.021)
$\beta$	-0.408 (0.104)	-0.423 (0.094)	-0.359 (0.071)	-0.096 (0.050)	0.037 (0.046)	-0.002 (0.031)	-0.270 (0.065)	-0.452 (0.076)	-0.250 (0.048)	-0.174 (0.037)	-0.083 (0.026)	-0.037 (0.022)
$\tau^2$	0.668 (0.136)	0.607 (0.144)	0.524 (0.151)	0.781 (0.097)	0.689 (0.110)	0.333 (0.228)	0.380 (0.202)	0.276 (0.282)	0.236 (0.322)	0.215 (0.352)	0.179 (0.424)	0.156 (0.486)

Table 2.4: Estimated values, their standard errors and t-values of  $\rho$  and estimated values of  $\alpha$  in the S-ARCH model in ref2.1 applied to the residuals by fitting SAR models to log returns of land price data year by year.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
$\rho$	0.082	0.189	0.227	0.274	0.392	0.192	0.357	0.228	0.211	0.255	0.344	0.277
$se(\rho)$	0.083	0.079	0.077	0.075	0.069	0.079	0.071	0.077	0.078	0.076	0.072	0.075
$t(\rho)$	0.987	2.398	2.925	3.633	5.672	2.435	5.018	2.954	2.708	3.350	4.788	3.687
$\alpha$	-0.254	-0.137	-0.227	0.192	0.348	-0.608	-0.132	-0.781	-0.929	-0.885	-0.719	-0.939

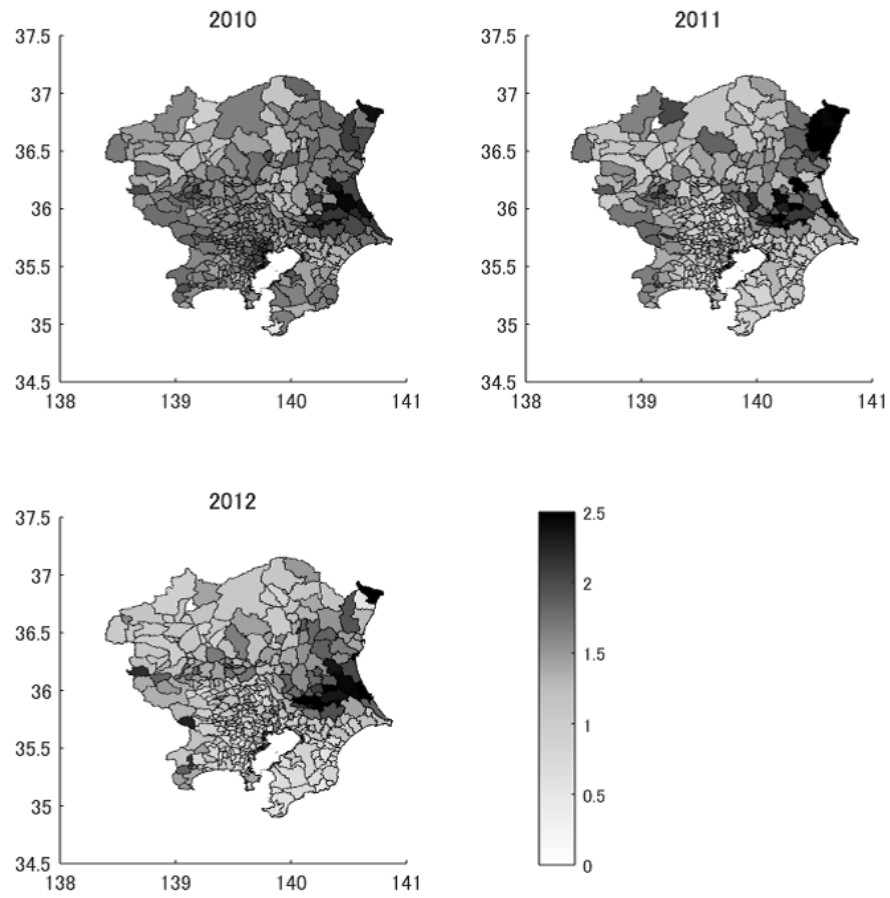


Figure 2.4: The identified volatilities in 2010-2012. Notice that the great earth quake occurred and the coastal areas were hit by great Tsunami in 2011.

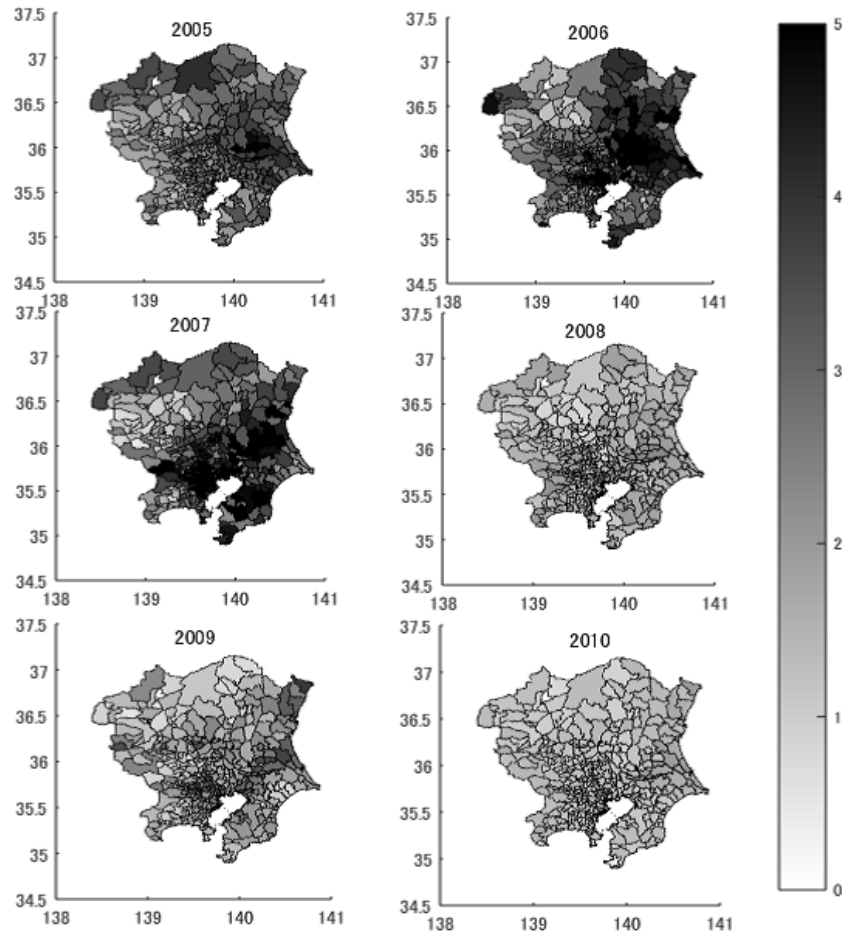


Figure 2.5: The identified volatilities in 2005-2010. Notice that Lehman shock occurred in 2008 when economic boom completed and serious recession began.



## Chapter 3

# Spatial GARCH Models

### Abstract

This paper extends a generalized autoregressive conditional heteroscedasticity (GARCH) model for time series to that for spatial data, which we call a spatial GARCH (S-GARCH) model. S-GARCH models specify conditional variances of spatial observations given those on surrounding areas, which is a good contrast with time series GARCH models that specify those given past observations. We employ a two step procedure to estimate parameters in S-GARCH models after transforming them into special kinds of spatial econometrics models, *ie* spatial autoregressive moving-average (SARMA) models, and derive consistency and asymptotic normality of the estimators. Simulation studies and applications to land price data in Tokyo areas are conducted to demonstrate empirical properties of S-GARCH models.

### 3.1 Introduction

Volatility models for financial time series have been developing with their fruitful applications in financial industries. The seminal work by Engle (1982) introduces a autoregressive conditional heteroscedasticity (ARCH) model, while Bollerslev (1986) extends it to a generalized ARCH (GARCH) model. These models have been widely used to identify volatilities that play significant roles in pricing options or measuring values at risk of financial positions in risk management. Subsequently, multivariate extensions have been conducted by Bollerslev et al (1988), Bollerslev (1990) and Engle and Kroner (1995) for modeling dynamic relationships of volatilities among multiple asset returns. A major challenge of multivariate volatility modeling is to overcome curse of dimensionality;  $n(n+1)/2$  components in volatility matrices for  $n$ -dimensional asset return series require complicated modeling and estimation procedures when  $n$  is large. One solution for the problem is to consider simpler structures of volatility matrices to reduce parameter dimensions.

Applications of spatial econometrics to volatility modeling have begun in recent years in order to extend volatility models for financial time series to those for spatial areal data. Spatial volatility is defined as a conditional variance on an area given observations over all other areas, which is a spatial analogy of time series volatility. Spatial weight matrix plays a key role to construct models for spatial volatilities. See Caporin and Paruolo (2008) and Borovkova and Lopuhaa (2012) for multivariate extensions of GARCH models by spatial econometrics methodology, Yan (2007) and Robinson (2009) for spatial extensions of stochastic volatility (SV) models, and Sato and Matsuda (2017) for those of ARCH models, which we named spatial ARCH (S-ARCH) models.

The aim of the paper is to conduct a spatial extension of GARCH models, which we call spatial generalized ARCH (S-GARCH) models. S-GARCH models are regarded as extensions of S-ARCH models of Sato and Matsuda (2017) in similar ways of extensions from time series ARCH to GARCH models. S-GARCH models are characterized by the following two features. First, spatial volatility at one area depends on volatilities as well as observations over surrounding neighbors, which is an extension of S-ARCH models whose volatility depends only on surrounding observations (Sato and Matsuda, 2017). Secondly, S-GARCH models can be transformed into a kind of spatial econometrics models called spatial autoregressive moving average (SARMA) models by which S-GARCH models are guaranteed to exist and several popular techniques in spatial econometrics literatures are to be applied to estimation and kriging for spatial volatilities.

For estimating parameters in S-GARCH models, we employ quasi-maximum likelihood (QML) estimation and prove consistency and asymptotic normality for QMLE. There are two kinds of popular estimation in spatial econometrics literatures. First one is estimation by moment methods. See Kelejian and Robinson (1993) and Kelejian and Prucha (1997, 1998) for two stage least squares, and Lee (2007) for the generalized method of moments (GMM) for spatial autoregressive (SAR) models and spatial autoregressive models with autoregressive disturbances (SARAR). In addition, see Dogan and Taspinar (2013) for GMM estimation for spatial autoregressive models with moving average disturbances (SARMA). The other one is estimation by QML methods. See Lee (2004), Yu et al (2008) and Su and Yang (2015) for QML estimation for SAR models and spatial dynamic panel (SDP) models and Yang (2018) for M-estimator based on QML for SDP models. This paper employs the latter approach for estimation, namely applies QML method to SARMA models to estimate parameters in S-GARCH models, as S-GARCH models can be transformed into SARMA models.

This paper proceeds as follows. Section 3.2 introduces S-GARCH models and discusses their characteristics. Section 3.3 proposes two step estimation procedure for S-GARCH model and derives asymptotic properties of the estimators. Section 3.4 demonstrates applications of S-GARCH models to simulated and real spatial data of land price data in Tokyo areas. Section 3.5 concludes the paper. All the proofs for the asymptotic properties in Section 3.3 are collected in the Appendix.

## 3.2 Spatial GARCH models

Suppose we observe spatial data  $y_i$  on a spatial area  $i$  for  $i = 1, \dots, n$ . We shall define S-GARCH models to describe spatial volatilities of  $y_i$  by

$$\begin{aligned} y_i &= \sqrt{h_i} \varepsilon_i, \\ \log h_i &= \lambda \sum_{j=1}^n w_{ij} \log h_j + \rho \sum_{j=1}^n w_{ij} \log y_j^2 + \alpha + \mathbf{z}_i' \delta, \end{aligned} \quad (3.1)$$

where  $\sqrt{h_i}$  is volatility,  $\varepsilon_i$  is an independent and identically distributed (i.i.d) random variable with mean zero and variance 1,  $\mathbf{z}_i$  is  $(k \times 1)$  non-stochastic regressors, and  $w_{ij}$  is a spatial weight that quantifies vicinity from area  $i$  to area  $j$ . The matrix  $W_n$ ,  $n$  by  $n$  matrix composed of  $w_{ij}$ , is called a spatial weight matrix. A spatial weight matrix is usually determined by geographical information of spatial data. The first order contiguity weight matrix is a standard choice for it (Sato and Matsuda, 2017). For parameters  $(\lambda, \rho, \alpha, \delta)'$  in this model,  $\lambda$  and  $\rho$  describes spatial interactions of volatilities and logged observations. S-GARCH models reduce to S-ARCH models proposed by Sato and Matsuda (2017) when  $\lambda$  is equal to 0.

S-GARCH models describe logged spatial volatilities by linear combinations of spatial lagged values of logged volatilities and observations, which is a spatial analogy of time series GARCH models. Although spatial and time series volatilities are completely different, it will be shown in Section 4 that they share some common features such as volatility clustering, namely a large change at one area causes large changes at surrounding areas.

We must notify here that S-GARCH models describe logged volatilities, while time series GARCH models describe volatilities without log transformation. The logged volatility makes it possible to re-express S-GARCH models as SARMA models. The SARMA expression results in providing two useful key features for S-GARCH models. First one is that existence conditions for S-GARCH models are easily checked by regarding them as SARMA models. Time series GARCH models are guaranteed to exist by regarding them as Markov processes (Fan and Yao, 2003). The same technique of Markov processes cannot be applied to S-GARCH models as there are no orders for spatial data unlike time series. The other one is that several popular tools such developed in spatial econometrics literatures can be employed for S-GARCH modeling. Quasi maximum likelihood (QML) method, one of critical estimation tools in spatial econometrics, shall be applied to estimate parameters in S-GARCH models in Section 3.

Let us re-express S-GARCH models as SARMA models. Denoting  $\log \mathbf{y}^2 = (\log y_1^2, \dots, \log y_n^2)'$ ,  $\log \mathbf{h} = (\log h_1, \dots, \log h_n)'$ ,  $\log \boldsymbol{\varepsilon}^2 = (\log \varepsilon_1^2, \dots, \log \varepsilon_n^2)'$ ,  $Z_n = (\mathbf{z}_1, \dots, \mathbf{z}_n)'$ ,  $\mathbf{1}_n = (1, \dots, 1)'$  and  $I_n$  is  $n \times n$  identity matrix, we have the vector representation for S-GARCH models,

$$\log \mathbf{y}^2 = \log \mathbf{h} + \log \boldsymbol{\varepsilon}^2 \quad (3.2)$$

$$\log \mathbf{h} = \lambda W_n \log \mathbf{h} + \rho W_n \log \mathbf{y}^2 + \alpha \mathbf{1}_n + Z_n \delta, \quad (3.3)$$

where  $W_n$  is the spatial weight matrix composed of  $w_{ij}$ . It is easy to see from (3.3) that,

$$\log \mathbf{h} = (I_n - \lambda W_n)^{-1}(\rho W_n \log \mathbf{y}^2 + \alpha \mathbf{1}_n + Z_n \delta),$$

Substituting the equation into (3.2), we obtain

$$\log \mathbf{y}^2 = (I_n - \lambda W_n)^{-1}(\rho W_n \log \mathbf{y}^2 + \alpha \mathbf{1}_n + Z_n \delta) + \log \boldsymbol{\varepsilon}^2,$$

Multiplying  $I_n - \lambda W_n$  from the left, we have

$$(I_n - \lambda W_n) \log \mathbf{y}^2 = \rho W_n \log \mathbf{y}^2 + \alpha \mathbf{1}_n + Z_n \delta + (I_n - \lambda W_n) \log \boldsymbol{\varepsilon}^2,$$

which reduces to

$$\log \mathbf{y}^2 = (\lambda + \rho) W_n \log \mathbf{y}^2 + \alpha \mathbf{1}_n + Z_n \delta + (I_n - \lambda W_n) \log \boldsymbol{\varepsilon}^2, \quad (3.4)$$

which is a SARMA model, one of popular spatial econometrics models (Dogan and Taspinar (2013)). It has been known that SARMA is guaranteed to exist when  $|\lambda| + |\rho| < 1$ , and hence S-GARCH models are guaranteed exist under the same condition as a result.

### 3.3 Estimation

We shall propose estimation of the parameters  $(\lambda, \rho, \alpha, \delta)'$  in S-GARCH models and derive the asymptotic properties of the estimators. Parameters are estimated by a two step procedure. First step is the estimation of  $(\lambda, \rho, \delta)'$  by QML method. The constant term  $\alpha$  shall be estimated separately in the second step, as  $\log \boldsymbol{\varepsilon}^2$  in (3.4) is not zero mean. and the estimator for constant term in the first step is biased. In second step,  $\alpha$  is estimated by the likelihood different from the one in the first step.

#### 3.3.1 First step estimation

Parameters  $\lambda, \rho$  and  $\delta$  are estimated in first step by the QML estimation method.

To apply QML method, we need to demean the error term. Observing in (3.4) that

$$\begin{aligned} \alpha \mathbf{1}_n + (I_n - \lambda W_n) \log \boldsymbol{\varepsilon}^2 &= \alpha \mathbf{1}_n + (I_n - \lambda W_n) \{ \log \boldsymbol{\varepsilon}^2 - E(\log \varepsilon_1^2) \mathbf{1}_n + E(\log \varepsilon_1^2) \mathbf{1}_n \}, \\ &= \{ \alpha + (1 - \lambda) E(\log \varepsilon_1^2) \} \mathbf{1}_n + (I_n - \lambda W_n) \{ \log \boldsymbol{\varepsilon}^2 - E(\log \varepsilon_1^2) \mathbf{1}_n \}, \end{aligned}$$

we see the intercept term has the bias by  $(1-\lambda)E(\log \varepsilon_1^2)$ . Denoting  $Y_n = \log \mathbf{y}^2$ ,  $X_n = [\mathbf{1}_n, \mathbf{Z}]$ ,  $V_n = \{\log \varepsilon^2 - E(\log \varepsilon_1^2) \mathbf{1}_n\}$  and  $\beta = (\{\alpha + (1-\lambda)E(\log \varepsilon_1^2)\}, \delta')'$ , we have the following modified representation,

$$Y_n = \lambda W_n Y_n + \rho W_n Y_n + X_n \beta + (I_n - \lambda W_n) V_n, \quad (3.5)$$

where  $V_n$  is the zero mean processes.

Now, let us consider the QML estimation of (3.5). Regarding  $v_i$ 's as independent Gaussian variables with mean zero and variance  $\sigma^2$ , the likelihood function of (3.5) is

$$\log L_n(\psi) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{V_n'(\theta, \beta) V_n(\theta, \beta)}{2\sigma^2} - \log |R_n(\lambda)| + \log |S_n(\theta)|, \quad (3.6)$$

where  $\theta = (\lambda, \rho)'$ ,  $\psi = (\beta', \sigma_2, \theta')'$ ,  $R_n(\lambda) = I_n - \lambda W_n$ ,  $R_n = I_n - \lambda_0 W_n$ ,  $S_n(\theta) = I_n - \lambda W_n - \rho W_n$ ,  $S_n = I_n - \lambda_0 W_n - \rho_0 W_n$  and  $V_n(\theta, \beta) = R_n^{-1}(\lambda)[S_n(\theta)Y_n - X_n\beta]$ . The QML estimator is the extreme estimator derived from the maximization of (3.6).

It is convenient to work with the concentrated likelihood by concentrating  $\beta$  and  $\sigma^2$  out for computation and asymptotic analysis on the estimator. From the first order condition of (3.6), the concentrated QML estimators of  $\beta$  and  $\sigma^2$  is

$$\begin{aligned} \hat{\beta}_n(\theta) &= (X_n' R_n'^{-1}(\lambda) R_n^{-1} X_n)^{-1} X_n' R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) Y_n, \\ \hat{\sigma}_n^2(\theta) &= \frac{\hat{V}_n'(\theta) \hat{V}_n(\theta)}{n}, \end{aligned}$$

where  $\hat{V}_n(\theta) = R_n^{-1}(\lambda)[S_n(\theta)Y_n - X_n\hat{\beta}_n(\theta)]$ . The concentrated likelihood function of  $\theta$  is

$$\log L_n(\theta) = -\frac{n}{2}(\log(2\pi) + 1) - \frac{n}{2} \log \hat{\sigma}_n^2(\theta) - \log |R_n(\lambda)| + \log |S_n(\theta)| \quad (3.7)$$

The QML estimator  $\hat{\theta}_n$  maximizes the concentrated likelihood function (3.7) and the QML estimators of  $\beta$  and  $\sigma^2$  are  $\hat{\beta}_n(\hat{\theta}_n)$  and  $\hat{\sigma}_n^2(\hat{\theta}_n)$ , respectively.

For our analysis of the asymptotic properties of first step estimators, we need the following assumptions:

**Assumption 6.** The disturbances  $\{v_i\}, i = 1, \dots, n$  are i.i.d. across  $i$  with zero mean, variance  $\sigma_0^2$  and  $E|v_i|^{4+\eta} < \infty$  for some  $\eta > 0$ .

**Assumption 7.** The elements  $w_{n,ij}$  of  $W_n$  are nonnegative and row normalized and the column sums of  $W_n$  are uniformly bounded.

**Assumption 8.** The space  $\Theta$  is compact, and the true parameter  $\theta_0$  lies in its interior.

**Assumption 9.** The matrix  $S_n, S_n(\theta), R_n$ , and  $R_n(\lambda)$  are uniformly bounded both row and column sums and nonsingular.

**Assumption 10.** The elements of  $X_n$  are uniformly bounded constants. The  $\lim_{n \rightarrow \infty} \frac{1}{n} (X'_n R_n'^{-1}(\lambda) R_n^{-1}(\lambda) X_n)$  exists and is nonsingular.

**Assumption 11.**  $0 \leq \underline{c}_y \leq \inf_{\theta \in \Theta} \gamma_{\min}(\text{Var}(S_n(\theta)Y_n)) \leq \sup_{\theta \in \Theta} \gamma_{\max}(\text{Var}(S_n(\theta)Y_n)) \leq \bar{c}_y < \infty$ .

**Assumption 12.**  $0 \leq \underline{c}_r \leq \inf_{\lambda \in \Lambda} \gamma_{\min}(R_n'^{-1}(\lambda) R_n^{-1}(\lambda)) \leq \sup_{\lambda \in \Lambda} \gamma_{\max}(R_n'^{-1}(\lambda) R_n^{-1}(\lambda)) \leq \bar{c}_r < \infty$

**Assumption 13.**  $\lim_{n \rightarrow \infty} \frac{1}{n} \beta'_0 X'_n S_n'^{-1} S'_n(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0 \neq 0$ ,  
where  $M_n = I_n - R_n^{-1} X_n (X'_n R_n'^{-1}(\lambda) R_n^{-1}(\lambda) X_n)^{-1} X'_n R_n'^{-1}$ .

To derive the consistency of the QML estimators, we need to show the identification of  $\theta_0$ . Define  $Q_n(\theta) = \max_{\beta, \sigma^2} E(\log L_n(\psi))$ . The optimal solutions of this maximization problem are given by

$$\begin{aligned} \beta_n^*(\theta) &= (X'_n R_n'^{-1}(\lambda) R_n^{-1} X_n)^{-1} X'_n R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) E(Y_n), \\ \sigma_n^{2*} &= \frac{1}{n} E(V_n'^*(\theta) V_n^*(\theta)), \end{aligned}$$

where  $V_n^*(\theta) = R_n^{-1}(\lambda) [S_n(\theta) Y_n - X_n \beta_n^*(\theta)]$ . Therefore,

$$Q_n(\theta) = -\frac{n}{2} (\log(2\pi) + 1) - \frac{n}{2} \log \sigma_n^{2*}(\theta) - \log |R_n(\lambda)| + \log |S_n(\theta)|,$$

and identification of  $\theta_0$  is based on  $\frac{1}{n} Q_n(\theta)$ .

Consistency of the QML estimators  $\hat{\theta}$  follows from the uniform convergence of  $\frac{1}{n} \log L_n(\theta) - \frac{1}{n} Q_n(\theta)$  to zero on  $\Theta$  and identification of  $\theta_0$ .

**Theorem 3.** Under Assumptions 6-13,  $\theta_n$  is a consistent estimator of  $\theta_0$ .

To derive the asymptotic distribution of the QMLE  $\hat{\psi}_n$ , we need to make the Taylor expansion of  $\frac{\partial}{\partial \psi} \log L_n(\hat{\psi}) = 0$  at  $\psi_0$ . The first-order derivatives of the log-likelihood function at  $\psi_0$  are

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \beta} &= \frac{1}{\sigma_0^2 \sqrt{n}} X'_n R_n'^{-1} V_n, \\ \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \sigma^2} &= \frac{1}{2\sigma_0^4 \sqrt{n}} (V_n' V_n - n\sigma_0^2), \\ \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \rho} &= \frac{1}{\sigma_0^2 \sqrt{n}} \beta'_0 X'_n S_n'^{-1} W_n' R_n'^{-1} V_n + \frac{1}{\sigma_0^2 \sqrt{n}} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} V_n - \sigma_0^2 \text{tr}(S_n^{-1} W_n)), \\ \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \lambda} &= \frac{1}{\sigma_0^2 \sqrt{n}} \beta'_0 X'_n S_n'^{-1} W_n' R_n'^{-1} V_n + \\ &\quad \frac{1}{\sigma_0^2 \sqrt{n}} (V_n' (R_n' S_n'^{-1} W_n' R_n'^{-1} - W_n' R_n'^{-1}) V_n - \sigma_0^2 \text{tr}(S_n^{-1} W_n) + \sigma_0^2 \text{tr}(R_n^{-1} W_n)), \end{aligned}$$

where  $\text{tr}(\cdot)$  denote the trace of a matrix.

These involve linear and quadratic function of  $V_n$ . The asymptotic distribution of these score functions are derived from the central limit theorems for linear-quadratic forms in Kelejian and Prucha (2001).

The variance matrix of  $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \psi} \log L_n(\psi_0)$  is

$$E\left(\frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \psi} \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \psi'}\right) = -E\left(\frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'}\right) + \Omega_{\psi,n},$$

where  $-E(\frac{1}{n} \frac{\partial^2}{\partial \psi \partial \psi'} \log L_n(\psi_0))$  is the average Hessian matrix and  $\Omega_{\psi,n}$  is a symmetric matrix and both are given in Appendix A. When  $V_n$  is normally distributed,  $\Omega_{\psi,n} = 0$ .

The score function and Hessian matrix have proper asymptotic behavior, therefore we have the following theorem.

**Theorem 4.** *Under Assumptions 6-13,*

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} N(0, \Sigma_\psi^{-1} + \Sigma_\psi^{-1} \Omega_\psi \Sigma_\psi^{-1}),$$

where  $\Sigma_\psi = -\lim_{n \rightarrow \infty} E(\frac{1}{n} \frac{\partial^2}{\partial \psi \partial \psi'} \log L_n(\psi_0))$  and  $\Omega_\psi = \lim_{n \rightarrow \infty} \Omega_{\psi,n}$ .  $\Sigma_\psi$  and  $\Omega_\psi$  assume to exist and  $-\Sigma_\psi$  to be positive definite, sufficiently large  $n$ . When errors are normally distributed,  $\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} N(0, \Sigma_\psi^{-1})$ .

### 3.3.2 Second step estimation

We move to estimation of  $\alpha$  in (3.4). As the estimated constant in the first step is biased, we need to estimate  $\alpha$  separately in the second step. In the second step, we employ the quasi likelihood by regarding  $\varepsilon_i$ , not  $\log \varepsilon_i^2$ , as Gaussian. Re-expressing (3.4) as

$$\begin{aligned} U_n(\phi) &= R_n^{-1}(\lambda)(S(\theta)Y - \alpha \mathbf{1}_n - Z_n \delta), \\ &= R_n^{-1}(\lambda)(S(\theta)Y - Z_n \delta) - \frac{\alpha}{1 - \lambda} \mathbf{1}_n, \\ &= C - \frac{\alpha}{1 - \lambda} \mathbf{1}_n, \end{aligned}$$

where  $C = R_n^{-1}(\lambda)(S(\theta)Y - Z_n \delta)$ , and seeing that the probability density function of  $\log \varepsilon^2$  for a standard normal variable  $\varepsilon$  is shown by Lee (2012, p. 379) as

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \exp(x) + \frac{1}{2} x\right), \quad (3.8)$$

we obtain the quasi likelihood function for  $\phi$  as

$$\log L_n(\phi) = \frac{n}{2} \log 2\pi - \sum_{i=1}^n \left\{ -\frac{1}{2} \frac{\exp(C_i)}{\exp(\frac{\alpha}{1-\lambda})} + \frac{1}{2} \left( C_i - \frac{\alpha}{1-\lambda} \right) \right\} - \log |R_n(\lambda)| + \log |S_n(\theta)|,$$

where  $C_i$  is the  $i$ -th element of  $C$ . Differentiating the likelihood with respect to  $\alpha$  and solving the equation by setting it to be 0, we have the relation

$$\alpha_n(\lambda, \rho, \delta) = (1 - \lambda) \log \left( \frac{1}{n} \sum_{i=1}^n \exp(C_i) \right).$$

Replacing  $(\lambda, \rho, \delta)'$  with the QML estimator  $(\hat{\lambda}, \hat{\rho}, \hat{\delta})'$  in the first step, we finally obtain

$$\hat{\alpha}_n = (1 - \hat{\lambda}) \log \left( \frac{1}{n} \sum_{i=1}^n \exp \left[ \{ R_n^{-1}(\hat{\lambda})(S(\hat{\theta})Y_n - Z_n\hat{\delta}) \}_i \right] \right), \quad (3.9)$$

as an estimator for  $\alpha$ .

The estimator  $\hat{\alpha}_n$  has consistency.

**Theorem 5.** *Under Assumptions 6-13,  $\hat{\alpha}_n$  is a consistent estimator of  $\alpha_0$ .*

## 3.4 Empirical analysis

We examine empirical properties of S-GARCH models by applying to simulated and land price data in Tokyo areas. Monte Carlo experiments are carried out to investigate finite sample performances of the two stage estimators, while land price data in Tokyo is employed to demonstrate practical performances of volatilities identified by S-GARCH models.

### 3.4.1 Simulation studies

To investigate finite sample properties of the two stage estimators, we simulate areal data by S-GARCH model in (3.1), where  $z_i$ 's are randomly generated from independent normal distributions and the spatial weights matrix is generated according to Rook contiguity and row normalizing. For the error terms,  $\varepsilon_i$ , we consider the three cases: (i) standard normal distributions, (ii) chi-squared distributions with 3 degrees of freedom and (iii) log normal distributions. For  $\phi = (\lambda, \rho, \alpha, \beta)'$ , we examine the three cases of  $(0.9, 0.05, 0.5, 1)$ ,  $(0.45, 0.45, 0.5, 1)$ ,  $(0.05, 0.9, 0.5, 1)'$  when the sample size  $n$  is 100 and 400. Each set of Monte Carlo results is based on 1000 repetitions of the two step estimation.

The empirical means and square root of mean squared errors (RMSE) for the two stage estimators are reported in Table 3.1. The results show the estimators in the first step,  $(\hat{\lambda}, \hat{\rho}, \hat{\beta})'$  are nearly unbiased and not sensitive to the choice of the error distributions. On the other hand, the second step estimator,  $\hat{\alpha}$  depends on the error distribution, ie, is more biased and has larger RMSE for more deviations from Gaussian.

### 3.4.2 Land price data analysis

We shall apply S-GARCH models to land price data in Tokyo area in order to demonstrate identification of spatial volatilities.



Let us introduce land price data used in this section. We use prefectural land price research as land price data. The Japanese Ministry of Land, Infrastructure, Transport, and Tourism publishes land prices on sampling points scattered irregularly all over Japan in the form of price per  $m^2$  in July. We focus on the land prices over Tokyo area (Tokyo, Kanagawa, Saitama, Chiba, Tochigi, Ibaraki, Gunma) observed yearly from 2009 to 2014. Averaging log returns of land price in municipal units, we obtain land price as yearly observations of areal data. Namely, land price data consists of yearly averaged log returns over 344 discrete areal unit's from 2010 to 2014.

Before application of S-GARCH models, we remove spatial correlations by fitting spatial autoregressive (SAR) models year by year. This is a spatial analogy of ARMA model fitting before fitting GARCH models to remove temporal correlations. A SAR model is

$$y_i = \zeta + \kappa \sum_{j=1}^{344} w_{ij} y_j + u_{it}, u_{it} \sim i.i.d(0, \tau^2).$$

where  $W = (w_{ij})$  is the first-order contiguity relation that takes 1 when two units share a common boarder.

We apply S-GARCH models to the residuals obtained after fitting SAR models year by year, where the spatial weight matrix in the S-GARCH model fitting is designed as the same as that in the SAR model fitting. As an explanatory variable for the regression term, we employ the logged area size of each municipal unit, which works to guarantee the identification condition in Assumption 13. Table 3.2 shows the estimated values of  $\lambda, \rho, \alpha$  and  $\beta$ , where  $\alpha$  and  $\beta$  are the intercept and regression coefficient, respectively. The standard errors of  $\hat{\lambda}$  and  $\hat{\rho}$  are evaluated by replacing the population moments with the corresponding sample moments in Theorem 3. Figure 3.1 expresses the spatial volatility identified by

$$\log \hat{\mathbf{h}} = (I_n - \hat{\lambda} W_n)^{-1} (\hat{\rho} W_n \log \mathbf{y}^2 + \hat{\alpha} \mathbf{1}_n + \mathbf{x} \hat{\beta}),$$

where  $\mathbf{x}$  is the independent variable of logged area size..

we find that  $\hat{\lambda}$ , the strength of interactions among spatial volatilities, are significant after the Great East Japan Earthquake in 2011 until 2013 from Table 3.2. This suggests that spatial volatility in land prices may have strengthened when the big event occurs.  $\hat{\rho}$ , the strength between spatial volatility and logged squared observations, is as large as the estimator for that of time series GARCH models. It is seen that  $\hat{\lambda} + \hat{\rho}$  is estimated to be close to 1 between 2011 and 2013, which likely causes volatility clustering around coastal areas. From Figure 3.1, we find that volatilities not only at coastal areas hit by the Tsunami but also areas near Fukushima are identified to be high and , which suggests the effects of Fukushima nuclear accidents. Finally in comparison between identified volatilities by fitting S-ARCH and S-GARCH models in (3.2), we observe that S-GARCH fits better in terms of AIC with global spillover, which means the identified volatility by S-GARCH models are more highly spatially correlated

than those of S-ARCH models. From figure 3.2, we find global spillover effects of volatility in land price data.

### 3.5 Conclusion

We have proposed a spatial generalized autoregressive conditional heteroskedasticity (S-GARCH) model as extension of a spatial autoregressive conditional heteroskedasticity (S-ARCH) model by Sato and Matsuda (2017). By re-expressing S-GARCH as spatial autoregressive moving average (SARMA) models, we employ spatial econometrics methodology to estimate the parameters by the two step procedure, and establish rigorous asymptotic results. Applications to land price data in Tokyo demonstrate that S-GARCH models detect several interesting features of spatial volatilities caused by the Great East Japan Earthquake in 2011.

Finally let us introduce possible extensions S-GARCH models. We employed the first-order contiguity relations to construct a spatial weight matrix, which is the simplest choice. It is desired to check what kind of spatial weight matrix can improve the fitting of S-GARCH models. Spatio-temporal extension of the S-GARCH models are surely our next target that can provide much better ways for land price data analysis than the year by year fitting of S-GARCH models in this paper.

Table 3.1: The empirical means and root mean squared errors (RMSE) of the estimators.

$\phi$	normal			chi(3)			log normal		
	n=100			n=100			n=100		
	Bias	RMSE	n=400	Bias	RMSE	n=400	Bias	RMSE	n=400
0.9	0.029	0.082	0.007	0.029	0.029	0.007	0.032	0.078	0.009
0.05	-0.039	0.069	-0.009	0.026	0.040	0.066	-0.040	0.066	-0.010
0.5	0.039	0.378	0.006	0.105	0.015	0.310	0.015	0.310	-0.003
1.0	0.021	0.188	0.009	0.089	0.023	0.176	0.023	0.176	0.004
0.45	-0.060	0.238	-0.015	0.098	-0.065	0.243	-0.065	0.243	-0.015
0.45	-0.001	0.155	0.002	0.072	0.002	0.159	0.002	0.159	0.002
0.5	-0.014	0.292	-0.002	0.092	-0.054	0.313	-0.054	0.313	-0.007
1.0	0.034	0.232	0.012	0.113	0.044	0.229	0.044	0.229	0.018
0.05	-0.027	0.139	-0.011	0.080	-0.023	0.141	-0.023	0.141	-0.011
0.9	-0.011	0.108	0.002	0.069	-0.017	0.115	-0.017	0.115	0.002
0.5	-0.431	0.829	-0.100	0.240	-0.627	1.089	-0.627	1.089	-0.114
1.0	0.013	0.236	0.007	0.114	0.022	0.228	0.022	0.228	0.006

Note:  $\phi = (\lambda, \rho, \alpha, \beta)'$

Table 3.2: Estimated values and standard errors of  $\lambda$ ,  $\rho$ ,  $\alpha$  and  $\beta$  in S-ARCH and S-GARCH models, which are applied year by year to the residuals by fitting SAR models to land priced data.

	S-ARCH					S-GARCH				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
$\hat{\lambda}$						0.772	0.845	0.874	0.893	0.601
$se(\lambda)$						0.206	0.139	0.128	0.100	0.415
$\hat{\rho}$	0.240	0.244	0.274	0.279	0.184	0.110	0.076	0.059	0.060	0.104
$se(\rho)$	0.083	0.081	0.082	0.083	0.084	0.077	0.055	0.048	0.045	0.086
$\hat{\alpha}$	0.569	-0.518	-0.606	-0.193	-0.804	0.162	-0.121	-0.130	-0.021	-0.412
$\hat{\beta}$	-0.022	0.212	0.232	0.109	0.225	-0.001	0.052	0.049	0.025	0.120
AIC	1538.7	1481.7	1549.8	1573.8	1537.7	1536.6	1475.3	1547.9	1570.4	1537.9

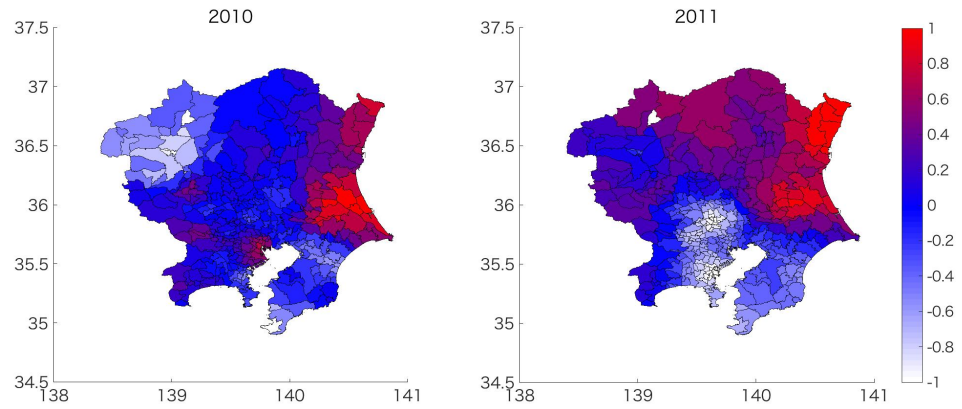


Figure 3.1: The identified volatilities in 2010 and 2011. The great earth quake occurred in 2011.

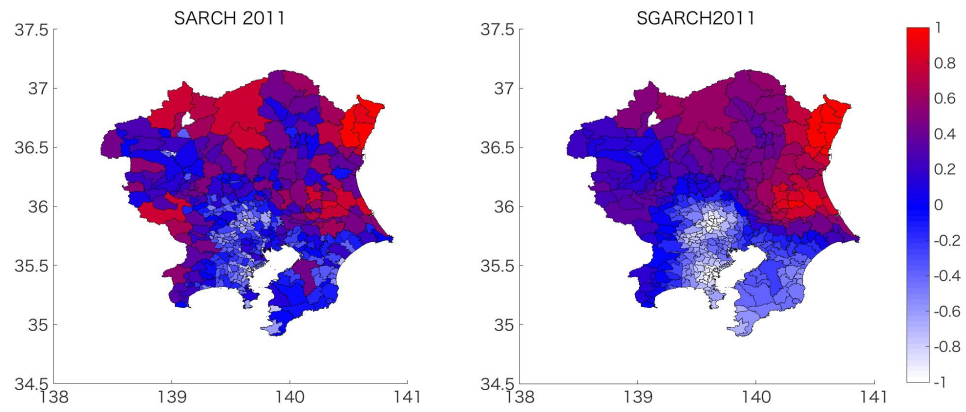


Figure 3.2: A comparison between identified volatilities by S-ARCH and S-GARCH models.

## A. Hessian, average Hessian and symmetric matrix $\Omega_{\psi,n}$

The Hessian matrix  $H_n(\psi) \equiv \frac{\partial^2}{\partial \psi \partial \psi'} \log L_n(\psi)$  has the elements:

$$\begin{aligned}
H_{\beta\beta'} &= -\frac{1}{\sigma^2} X'_n R_n'^{-1}(\lambda) R_n^{-1}(\lambda) X_n, \\
H_{\beta\sigma^2} &= -\frac{1}{\sigma^4} X'_n R_n'^{-1}(\lambda) V(\theta), \\
H_{\beta\rho} &= -\frac{1}{\sigma^2} X'_n R_n'^{-1}(\lambda) R_n^{-1}(\lambda) W_n Y_n, \\
H_{\beta\lambda} &= \frac{1}{\sigma^2} X'_n R_n'^{-1}(\lambda) (W'_n R_n'^{-1}(\lambda) V_n(\theta) + R_n^{-1}(\lambda) W_n V_n(\theta) - R_n^{-1}(\lambda) W_n Y_n), \\
H_{\sigma^2\sigma^2} &= \frac{n}{2\sigma^4} - \frac{V'_n(\theta) V_n(\theta)}{\sigma^6}, \\
H_{\sigma^2\rho} &= -\frac{1}{\sigma^4} Y'_n W'_n R_n'^{-1}(\lambda) V(\theta), \\
H_{\sigma^2\lambda} &= \frac{1}{\sigma^4} (V'_n(\theta) - Y'_n) W'_n R_n'^{-1}(\lambda) V_n(\theta), \\
H_{\rho\rho} &= -\frac{1}{\sigma^2} Y'_n W'_n R_n'^{-1}(\lambda) R_n^{-1}(\lambda) W_n Y_n - \text{tr}(S_n^{-1}(\theta) W_n S_n^{-1}(\theta) W_n), \\
H_{\rho\lambda} &= \frac{1}{\sigma^2} Y'_n W'_n R_n'^{-1}(\lambda) (W'_n R_n'^{-1}(\lambda) V(\theta) + R_n^{-1}(\lambda) W_n V_n(\theta) - R_n^{-1}(\lambda) W_n Y_n) \\
&\quad - \text{tr}(S_n^{-1}(\theta) W_n S_n^{-1}(\theta) W_n), \\
H_{\lambda\lambda} &= \frac{1}{\sigma^2} (Y'_n - V'_n(\theta)) W'_n R_n'^{-1}(\lambda) (2W'_n R_n'^{-1}(\lambda) V_n(\theta) + R_n^{-1}(\lambda) W_n V_n(\theta) - R_n^{-1}(\lambda) W_n Y_n) \\
&\quad + \text{tr}(R_n^{-1}(\lambda) W_n R_n^{-1}(\lambda) W_n) - \text{tr}(S_n^{-1}(\theta) W_n S_n^{-1}(\theta) W_n).
\end{aligned}$$

The average Hessian matrix  $\Sigma_{\psi,n} \equiv -E\left(\frac{1}{n} \frac{\partial^2}{\partial \psi \partial \psi'} \log L_n(\psi_0)\right)$  has the ele-

ments:

$$\begin{aligned}
\Sigma_{\beta\beta'} &= \frac{1}{n\sigma_0^2} X_n' R_n'^{-1} R_n^{-1} X_n, \\
\Sigma_{\beta\sigma^2} &= 0, \\
\Sigma_{\beta\rho} &= \frac{1}{n\sigma_0^2} X_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} X_n \beta_0, \\
\Sigma_{\beta\lambda} &= \frac{1}{n\sigma_0^2} X_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} X_n \beta_0, \\
\Sigma_{\sigma^2\sigma^2} &= \frac{1}{2\sigma_0^4}, \\
\Sigma_{\sigma^2\rho} &= \frac{1}{n\sigma_0^2} \text{tr}(W_n S_n^{-1}), \\
\Sigma_{\sigma^2\lambda} &= \frac{1}{n\sigma_0^2} \text{tr}(W_n S_n^{-1} - W_n R_n^{-1}), \\
\Sigma_{\rho\rho} &= \frac{1}{n\sigma_0^2} \beta_0' X_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} X_n \beta_0 + \\
&\quad \frac{1}{n} \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n + S_n^{-1} W_n S_n^{-1} W_n), \\
\Sigma_{\rho\lambda} &= \frac{1}{n\sigma_0^2} \beta_0' X_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} X_n \beta_0 + \\
&\quad \frac{1}{n} \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n + S_n^{-1} W_n S_n^{-1} W_n) \\
&\quad - \frac{1}{n} \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n + S_n^{-1} W_n R_n^{-1} W_n), \\
\Sigma_{\lambda\lambda} &= \frac{1}{n\sigma_0^2} \beta_0' X_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} X_n \beta_0 + \\
&\quad \frac{1}{n} \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n + S_n^{-1} W_n S_n^{-1} W_n) \\
&\quad - \frac{2}{n} \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n + S_n^{-1} W_n R_n^{-1} W_n) + \frac{1}{n} \text{tr}(R_n^{-1} W_n R_n^{-1} W_n + W_n' R_n'^{-1} R_n^{-1} W_n).
\end{aligned}$$

The symmetric matrix  $\Omega_{\psi,n}$  has the elemetns:

$$\begin{aligned}
\Omega_{\beta\beta'} &= 0, \\
\Omega_{\beta\sigma^2} &= \frac{\mu_3}{2n\sigma_0^6} X_n' R_n'^{-1} \mathbf{1}_n, \\
\Omega_{\beta\rho} &= \frac{\mu_3}{n\sigma_0^4} \sum_i^n \{(R_n^{-1} X_n)_i\}' (R_n^{-1} W_n S_n^{-1} R_n)_{ii}, \\
\Omega_{\beta\lambda} &= \frac{\mu_3}{n\sigma_0^4} \sum_i^n \{(R_n^{-1} X_n)_i\}' (R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)_{ii}, \\
\Omega_{\sigma^2\sigma^2} &= \frac{\mu_4 - 3\sigma_0^4}{4\sigma_0^8}, \\
\Omega_{\sigma^2\rho} &= \frac{\mu_3}{2n\sigma_0^6} \beta_0' X_n' S_n'^{-1} W_n' R_n'^{-1} \mathbf{1}_n + \frac{\mu_4 - 3\sigma_0^4}{2n\sigma_0^6} \text{tr}(S_n^{-1} W_n), \\
\Omega_{\sigma^2\lambda} &= \frac{\mu_3}{2n\sigma_0^6} \beta_0' X_n' S_n'^{-1} W_n' R_n'^{-1} \mathbf{1}_n + \frac{\mu_4 - 3\sigma_0^4}{2n\sigma_0^6} \text{tr}(S_n^{-1} W_n - R_n^{-1} W_n), \\
\Omega_{\rho\rho} &= \frac{2\mu_3}{n\sigma_0^4} \sum_{i=1}^n (R_n^{-1} W_n S_n^{-1} X_n \beta_0)_i (R_n^{-1} W_n S_n^{-1} R_n)_{ii} + \frac{\mu_4 - 3\sigma_0^4}{n\sigma_0^4} \sum_{i=1}^n \{(R_n^{-1} W_n S_n^{-1} R_n)_{ii}\}^2, \\
\Omega_{\rho\lambda} &= \frac{\mu_3}{n\sigma_0^4} \sum_{i=1}^n (R_n^{-1} W_n S_n^{-1} X_n \beta_0)_i (2R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)_{ii} \\
&\quad + \frac{\mu_4 - 3\sigma_0^4}{n\sigma_0^4} \sum_{i=1}^n (R_n^{-1} W_n S_n^{-1} R_n)_{ii} (R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)_{ii}, \\
\Omega_{\lambda\lambda} &= \frac{2\mu_3}{n\sigma_0^4} \sum_{i=1}^n (R_n^{-1} W_n S_n^{-1} X_n \beta_0)_i (R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)_{ii} \\
&\quad + \frac{\mu_4 - 3\sigma_0^4}{n\sigma_0^4} \sum_{i=1}^n \{(R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)_{ii}\}^2,
\end{aligned}$$

where  $\mu_3$  and  $\mu_4$  are the third and fourth moments of  $v_i$ s, respectively,  $(R_n^{-1} X_n)_i$  is the  $i$ -th row of  $(R_n^{-1} X_n)$ ,  $(R_n^{-1} W_n S_n^{-1} X_n \beta_0)_i$  is the  $i$ -th element of  $(R_n^{-1} W_n S_n^{-1} X_n \beta_0)$  and  $(R_n^{-1} W_n S_n^{-1} R_n)_{ii}$ ,  $(R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)_{ii}$  and  $(2R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)_{ii}$  are the  $(i, j)$ th element of  $(R_n^{-1} W_n S_n^{-1} R_n)$ ,  $(R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)$  and  $(2R_n^{-1} W_n S_n^{-1} R_n - R_n^{-1} W_n)$ , respectively.

## B. Some useful Lemmas

**Lemma 3.5.1** (Proposition 8.4.13, Bernstein (2009)). *Let  $A$  and  $B$  be matrices. We use  $\gamma_{\max}$  and  $\gamma_{\min}$  to denote the largest and smallest eigenvalues of a matrix. If  $A$  is symmetric and  $B$  is positive semi definite, then*

$$\gamma_{\min}(A) \text{tr}(B) \leq \text{tr}(AB) \leq \gamma_{\max}(A) \text{tr}(B).$$



**Lemma 3.5.2** (Lee, 2002, p.256; Lee, 2004, p1918). *Let  $\{A_n\}$  and  $\{B_n\}$  be two sequences of  $n \times n$  matrices that are uniformly bounded in both row and column sums and the elements of an  $n \times n$  matrix  $\{C_n\}$  be  $O(1)$  uniformly. Then*

1. *the sequence  $\{A_n B_n\}$  are uniformly bounded in both row and column sums,*
2. *the elements of  $C_n B_n$  have the uniform order  $O(1)$ , and*
3. *the elements of  $A_n$  are uniformly bounded and  $\text{tr}(A_n) = O(n)$ .*

**Lemma 3.5.3** (Lee, 2004, p1918). *The elements, the  $v_i$ 's of  $V_n$  are assumed to be i.i.d. with zero mean and a finite variance and the fourth moment of the  $v$ 's is assumed to exist. Suppose that  $A_n$  is a square matrix with its column sums being uniformly bounded and elements of the  $n \times K$  matrix  $Z_n$  are uniformly bounded. Let  $\{B_n\}$  be uniformly bounded either in row or column sums and their elements  $b_{n,ij}$  have  $O(1)$  uniformly in  $i$  and  $j$ . Then*

1.  $\frac{1}{\sqrt{n}} Z_n' A_n V_n = O_p(1)$  and
2.  $\frac{1}{n} E(V_n' B_n V_n) = O(1)$  and  $\frac{1}{n} [V_n' B_n V_n - E(V_n' B_n V_n)] = o_p(1)$ .

## C. Proofs of Theorems 3-5

### Proof of Theorem 3

The consistency of  $\hat{\theta}$  will follow from the uniform convergence of  $\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))$  to zero on  $\Theta$  and the uniqueness identification condition that, for any  $\epsilon > 0$ ,  $\limsup_{n \rightarrow \infty} \max_{\theta \in N_\epsilon^c(\theta_0)} \frac{1}{n}(Q_n(\theta) - Q_n(\theta_0)) < 0$ , where  $N_\epsilon^c(\theta_0)$  is the complement of an open neighborhood of  $\theta_0$  in  $\Theta$  of diameter  $\epsilon$  (Theorem 3.4 of White (1994)).

**Proof of the uniform convergence of  $\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))$**

First, we shall prove the uniform convergence of  $\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))$  to zero on  $\Theta$ . The proof follows from:

- (a)  $\inf_{\theta \in \Theta} \sigma_n^{*2}(\theta)$  is bounded away from zero,
- (b)  $\sup_{\theta \in \Theta} |\hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta)| = o_p(1)$ ,
- (c)  $\sup_{\theta \in \Theta} |\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))| = o_p(1)$ .

**Proof of (a)** By the definition of  $V_n^*(\theta)$ ,

$$\begin{aligned}
 V_n^*(\theta) &= R_n^{-1}(\lambda)(S_n(\theta)Y_n - X_n\beta_n^*(\theta)), \\
 &= R_n^{-1}(\lambda)S_n(\theta)Y_n - R_n^{-1}X_n(X_n'R_n^{-1}(\lambda)R_n^{-1}(\lambda)X_n)^{-1}X_n'R_n^{-1}(\lambda)R_n^{-1}(\lambda)S_n(\theta)E(Y_n), \\
 &= R_n^{-1}(\lambda)S_n(\theta)Y_n - P_nR_n^{-1}(\lambda)S_n(\theta)E(Y_n), \\
 &= M_nR_n^{-1}(\lambda)S_n(\theta)Y_n + P_nR_n^{-1}(\lambda)S_n(\theta)(Y_n - E(Y_n)),
 \end{aligned}$$

where,  $P_n = R_n^{-1}X_n(X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n)^{-1}X_n'R_n'^{-1}$  and  $M_n = I_n - P_n$ .

From the orthogonality between the two symmetric idempotent matrices  $M_n$  and  $P_n$ , we have,

$$\begin{aligned}\sigma_n^{*2}(\theta) &= \frac{1}{n}E(V_n'^*(\theta)V_n^*(\theta)), \\ &= \frac{1}{n}E[Y_n'S_n'(\theta)R_n'^{-1}(\lambda)M_nR_n^{-1}(\lambda)S_n(\theta)Y_n + \\ &\quad (Y_n - E(Y_n))'S_n'(\theta)R_n'^{-1}(\lambda)P_nR_n^{-1}(\lambda)S_n(\theta)(Y_n - E(Y_n))], \\ &= \frac{1}{n}E(Y_n'S_n'(\theta)R_n'^{-1}(\lambda)M_nR_n^{-1}(\lambda)S_n(\theta)E(Y_n) + \frac{1}{n}tr(R_n'^{-1}(\lambda)R_n^{-1}(\lambda)Var(S_n(\theta)Y_n)).\end{aligned}$$

The matrix  $M_n$  is positive semi definite because  $M_n$  is a symmetric idempotent matrix (Lemma 14.2.14 of Harville (1997)). Thus, the first term is non-negative uniformly in  $\theta \in \Theta$ .

Because the matrix  $Var(S_n(\theta)Y_n)$  is symmetric and  $\gamma_{min}Var(S_n(\theta)Y_n) > 0$  from the assumption, the matrix is positive semi definite (Theorem 3.25 of Schott (2005)). By Lemma 3.5.1, the second term is

$$\begin{aligned}\frac{1}{n}tr(R_n'^{-1}(\lambda)R_n^{-1}(\lambda)Var(S_n(\theta)Y_n)) &\geq \frac{1}{n}\gamma_{min}(R_n'^{-1}(\lambda)R_n^{-1}(\lambda))tr(Var(S_n(\theta)Y_n)), \\ &\geq \frac{1}{n}\mathcal{C}_r\mathcal{C}_y, \\ &> 0, \text{ uniformly in } \theta \in \Theta.\end{aligned}$$

It follow that  $\inf_{\theta \in \Theta} \sigma_n^{*2}(\theta)$  is bounded away from zero.

**Proof of (b)** Noting that

$$\begin{aligned}\hat{V}_n(\theta) &= R_n^{-1}(\lambda)(S_n(\theta)Y_n - X_n\hat{\beta}_n(\theta)), \\ &= R_n^{-1}(\lambda)S_n(\theta)Y_n - R_n^{-1}X_n(X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n)^{-1}X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)S_n(\theta)Y_n, \\ &= M_nR_n^{-1}(\lambda)S_n(\theta)Y_n.\end{aligned}$$

Hence,

$$\begin{aligned}\hat{\sigma}_n^2(\theta) &= \frac{1}{n}\hat{V}_n'(\theta)\hat{V}_n(\theta), \\ &= \frac{1}{n}Y_n'S_n'(\theta)R_n'^{-1}(\lambda)M_nR_n^{-1}(\lambda)S_n(\theta)Y_n.\end{aligned}$$

It follows that

$$\begin{aligned}\hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta) &= \frac{1}{n}Y_n'S_n'(\theta)R_n'^{-1}(\lambda)M_nR_n^{-1}(\lambda)S_n(\theta)Y_n - \frac{1}{n}E(Y_n'S_n'(\theta)R_n'^{-1}(\lambda)M_nR_n^{-1}(\lambda)S_n(\theta)Y_n) \\ &\quad - \frac{1}{n}E((Y_n - E(Y_n))'S_n'(\theta)R_n'^{-1}(\lambda)P_nR_n^{-1}(\lambda)S_n(\theta)(Y_n - E(Y_n))), \\ &= (Q_1 - EQ_1) - EQ_2,\end{aligned}$$

where,  $Q_1 = \frac{1}{n} Y_n' S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) Y_n$  and  $EQ_2 = \frac{1}{n} E((Y_n - E(Y_n))' S_n'(\theta) R_n'^{-1}(\lambda) P_n R_n^{-1}(\lambda) S_n(\theta) (Y_n - E(Y_n)))$ .

To show the result, it sufficient to show  $Q_1 - EQ_1 \xrightarrow{p} 0$  and  $EQ_2 \rightarrow 0$ , uniformly in  $\theta \in \Theta$ .

First, we show that  $Q_1 - EQ_1 \xrightarrow{p} 0$  uniformly in  $\theta \in \Theta$ . By Theorem 1 of Andrews (1992), the uniform convergence of  $Q_1 - EQ_1$  to zero in probability follows from the pointwise convergence for each  $\theta \in \Theta$  and stochastic equicontinuity of  $Q_1$ , i.e., for any  $\epsilon > 0$ , there exists a positive number  $\delta$  such that  $\limsup_{n \rightarrow \infty} P(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} > \epsilon) < \epsilon$ , where  $B(\theta, \delta)$  denote a closed ball in  $\Theta$  of radius  $\delta \geq 0$  centered at  $\theta$ .

First of all, the pointwise convergence of  $Q_1 - EQ_1$  will be shown. We have, by the identity:  $Y_n = S_n^{-1} X_n \beta_0 + S_n^{-1} R_n V_n$ ,

$$\begin{aligned} Q_1 &= \frac{1}{n} (S_n^{-1} X_n \beta_0 + S_n^{-1} R_n V_n)' S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) (S_n^{-1} X_n \beta_0 + S_n^{-1} R_n V_n), \\ &= \frac{1}{n} (\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0 + \\ &\quad 2\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n V_n \\ &\quad + V_n' R_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n V_n), \\ &= Q_{1,1}(\theta) + 2Q_{1,2}(\theta) + Q_{1,3}(\theta), \end{aligned}$$

where  $Q_{1,1}(\theta) = \frac{1}{n} (\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0)$ ,  $Q_{1,2}(\theta) = \frac{1}{n} (\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n V_n)$  and  $Q_{1,3}(\theta) = \frac{1}{n} (V_n' R_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n V_n)$ . The two terms  $Q_{1,2}(\theta)$  and  $Q_{1,3}(\theta)$  are stochastic.

For the second term, the column sums of  $S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n$  are uniformly bounded from assumption 8 and Lemma 3.5.2 and  $E(Q_{1,2}(\theta)) = 0$ . Thus, the pointwise convergence of  $Q_{1,2}(\theta) - E(Q_{1,2}(\theta))$  follow from Lemma 3.5.3. Similarly, the column sums of  $R_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n$  are uniformly bounded and the pointwise convergence of  $Q_{1,3}(\theta) - E(Q_{1,3}(\theta))$  follows from Lemma 3.5.3. Therefore,  $Q_1 - EQ_1 \xrightarrow{p} 0$ , for each  $\theta \in \Theta$ .

Next, we show that  $Q_1$  is stochastic equicontinuous. We have by the mean value theorem:

$$\begin{aligned} Q_{1,\ell}(\theta_1) - Q_{1,\ell}(\theta_2) &= \frac{\partial}{\partial \theta'} Q_{1,\ell}(\bar{\theta})(\theta_2 - \theta_1), \\ &\leq \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta'} Q_{1,\ell}(\theta) \right| (\theta_2 - \theta_1), \end{aligned}$$

where  $\ell = 1, 2, 3$  and  $\bar{\theta}$  lies between  $\theta_1$  and  $\theta_2$ . For stochastic equicontinuous, it suffices to show that  $\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta'} Q_{1,\ell}(\theta) \right| = O_p(1)$  by Theorem 21.10 of Davidson (1994). Let  $\Pi_1$  be  $S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1}$ ,  $\Pi_2$  be  $\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n$  and  $\Pi_3$  be  $R_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n$ . The partial derivatives  $\frac{\partial}{\partial \theta'} \Pi_{1,\ell}$  take simple form and consequently  $\frac{\partial}{\partial \theta'} \Pi_{1,\ell}$  are also uniformly bounded in both row and column sums. For  $Q_{1,1}$ , for any  $\theta$ , the elements of  $\beta_0' X_n' \frac{\partial}{\partial \theta'} S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1}$

and  $X_n\beta_0$  are uniformly bounded. Thus, there exists constants  $c_1$  and  $c_2$  such that  $|\{\beta'_0 X'_n(\frac{\partial}{\partial\theta'} S_n'^{-1} S'_n(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1})\}_i| \leq c_1$  and  $|(X_n\beta_0)_i| \leq c_2$  where  $\{\beta'_0 X'_n(\frac{\partial}{\partial\theta'} S_n'^{-1} S'_n(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1})\}_i$  and  $(X_n\beta_0)_i$  are the  $i$ -th elements of each vector. It follows that  $|\frac{\partial}{\partial\theta'} Q_{1,1}| \leq c_1 c_2 = O(1)$ . For  $Q_{1,2}$ , for any  $\theta$ ,  $|\frac{\partial}{\partial\theta'} \Pi_{1,2,i}| \leq c_3$  where  $\frac{\partial}{\partial\theta'} \Pi_{1,2,i}$  is the  $i$ -th element of  $\frac{\partial}{\partial\theta'} \Pi_{1,2}$ . Therefore, from Lemma 3.5.3,  $P(|\frac{\partial}{\partial\theta'} Q_{1,2}| > M) \leq P(|\frac{1}{n} \sum_{i=1}^n c_3 v_i| > M) = O(n^{-\frac{1}{2}})$ . For  $Q_{1,3}$ , for any  $\theta$ ,  $|\frac{\partial}{\partial\theta'} \Pi_{1,3,ij}| \leq c_4$  where  $\frac{\partial}{\partial\theta'} \Pi_{1,3,ij}$  is the  $(i, j)$ th element of  $\frac{\partial}{\partial\theta'} \Pi_{1,3}$ . Thus, from Lemma 3.5.3,  $P(|\frac{\partial}{\partial\theta'} Q_{1,3}| > M) \leq P(|\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n c_4 v_i v_j| > M) = O(1)$ . Thus,  $\sup_{\theta \in \Theta} |\frac{\partial}{\partial\theta'} Q_{1,\ell}(\theta)| = O_p(1)$ . It follows that  $Q_1$  is stochastic equicontinuous. Hence, by Theorem 1 of Andrews (1992),  $Q_1 - EQ_1 \xrightarrow{p} 0$  uniformly in  $\theta \in \Theta$ .

Secondly, we show that  $EQ_2 \rightarrow 0$ , uniformly in  $\theta \in \Theta$ . There exist  $\underline{c}_x$  such that

$0 < \underline{c}_x \leq \inf_{\lambda \in \Lambda} \gamma_{\min}(\frac{1}{n} X'_n R_n'^{-1} R_n^{-1} X)$  from assumption. By Assumption, Lemma 3.5.1 and 3.5.2 and theorem 3.4 of Schott (2005), We have,

$$\begin{aligned}
EQ_2 &= \frac{1}{n} E((Y_n - E(Y_n))' S'_n(\theta) R_n'^{-1}(\lambda) P_n R_n^{-1}(\lambda) S_n(\theta) (Y_n - E(Y_n))), \\
&= \frac{1}{n} \text{tr}(R_n'^{-1}(\lambda) P_n R_n^{-1}(\lambda) \text{Var}(S_n(\theta) Y_n)), \\
&= \frac{1}{n} \text{tr}(R_n'^{-1}(\lambda) R_n^{-1} X_n (X'_n R_n'^{-1}(\lambda) R_n^{-1}(\lambda) X_n)^{-1} X'_n R_n'^{-1} R_n^{-1}(\lambda) \text{Var}(S_n(\theta) Y_n)), \\
&\leq \frac{1}{n} \gamma_{\min}^{-1}(X'_n R_n'^{-1} R_n^{-1} X) \gamma_{\max}^2(R_n'^{-1}(\lambda) R_n^{-1}(\lambda)) \gamma_{\max}(\text{Var}(S_n(\theta) Y)) \text{tr}(X'_n X_n), \\
&= \frac{1}{n} \gamma_{\min}^{-1} \left( \frac{X'_n R_n'^{-1} R_n^{-1} X}{n} \right) \gamma_{\max}^2(R_n'^{-1}(\lambda) R_n^{-1}(\lambda)) \gamma_{\max}(\text{Var}(S_n(\theta) Y)) \frac{1}{n} \text{tr}(X'_n X_n), \\
&\leq \frac{1}{n} \underline{c}_x^{-1} \bar{c}_r^2 \bar{c}_y \frac{1}{n} \text{tr}(X'_n X_n), \\
&= O(n^{-1})
\end{aligned}$$

Hence,  $EQ_2 \rightarrow 0$ , uniformly in  $\theta \in \Theta$ .

Therefore,  $\sup_{\theta \in \Theta} |\hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta)| = o_p(1)$ , completing the proof of (b).

**Proof of (C)** We show that  $\sup_{\theta \in \Theta} |\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))| = o_p(1)$ . Note that

$$\frac{1}{n}(\log L_n(\theta) - Q_n(\theta)) = -\frac{1}{2}(\log \hat{\sigma}_n^2(\theta) - \log \sigma_n^{*2}(\theta)).$$

By the Taylor expansion,

$$|\log \hat{\sigma}_n^2(\theta) - \log \sigma_n^{*2}(\theta)| = \frac{1}{\tilde{\sigma}_n^2(\theta)} |\hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta)|,$$

where  $\tilde{\sigma}_n^2(\theta)$  lies between  $\hat{\sigma}_n^2(\theta)$  and  $\sigma_n^{*2}(\theta)$ . From the proof (a) and (b), it follows that  $\hat{\sigma}_n^2(\theta)$  is uniformly bounded away from zero on  $\Theta$ . Moreover,  $\tilde{\sigma}_n^2(\theta)$  is also uniformly bounded away from zero on  $\Theta$  because  $\tilde{\sigma}_n^2(\theta)$  exists between  $\hat{\sigma}_n^2(\theta)$

and  $\sigma_n^{*2}(\theta)$  and thereby  $\frac{1}{\sigma_n^{*2}(\theta)}$  is uniformly bounded. As  $\hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta)$  converges in probability to zero uniformly on  $\Theta$ ,  $|\log \hat{\sigma}_n^2(\theta) - \log \sigma_n^{*2}(\theta)| = o_p(1)$  uniformly on  $\Theta$ .

Consequently,  $\sup_{\theta \in \Theta} \left| \frac{1}{n} (\log L_n(\theta) - Q_n(\theta)) \right| = o_p(1)$ .

### Proof of the identification uniqueness condition

Secondly, we shall prove the identification uniqueness condition. The proof follow from:

- (i)  $\frac{1}{n} Q_n(\theta)$  is uniformly equicontinuous on  $\Theta$ .
- (ii) Show some properties of an auxiliary model.
- (iii) Show that the identification uniqueness condition holds.

**Proof of (i)** We show that  $\frac{1}{n} Q_n(\theta) = \frac{1}{2}(\log 2\pi + 1) - \frac{1}{2} \log \sigma_n^{*2}(\theta) - \frac{1}{n} \log |R_n(\lambda)| + \frac{1}{n} \log |S_n(\theta)|$  is uniformly equicontinuous on  $\Theta$ . It is sufficient to show that partial derivatives of each term are uniformly bounded. The uniform continuity of  $\log \sigma_n^{*2}(\theta)$  on  $\Theta$  follows because  $\frac{1}{\sigma_n^{*2}(\theta)}$  is uniformly bounded since  $\sigma_n^{*2}(\theta)$  is uniformly bounded away from zero on  $\Theta$ . For  $\frac{1}{n} \log |R_n(\lambda)|$ ,  $\frac{d}{d\lambda} \frac{1}{n} \log |R_n(\lambda)| = \frac{1}{n} \text{tr}(R_n^{-1}(\lambda) W_n)$ . From assumption and Lemma 3.5.2, the elements of  $R_n^{-1}(\lambda) W_n$  are uniformly bounded. Thus,  $\frac{1}{n} \text{tr}(R_n^{-1}(\lambda) W_n) = O(1)$  from Lemma 3.5.2. Similarly,  $\frac{\partial}{\partial \theta} \frac{1}{n} \log |S_n(\theta)| = O(1)$ . Hence,  $\frac{1}{n} Q_n(\theta)$  is uniformly equicontinuous on  $\Theta$ .

**Proof of (ii)** It is useful to establish an auxiliary process:

$$Y_n = \lambda W_n Y_n + \rho W_n Y_n + R_n(\lambda) V_n,$$

where  $V_n \sim N(0, \sigma_0^2 I_n)$ . The log-likelihood function of the above auxiliary process is given by

$$\begin{aligned} \log L_{p,n}(\theta, \sigma^2) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2(\theta)) - \log |R_n(\lambda)| + \log |S_n(\theta)| \\ &\quad - \frac{1}{2\sigma^2} Y_n' S_n'(\theta) R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) Y_n. \end{aligned}$$

Let  $E_p$  be the expectation under this auxiliary process. Define  $Q_{p,n}(\theta) = \max_{\sigma^2} E_p(\log L_{p,n}(\theta))$ . The optimal solutions of this maximization problem is

$$\begin{aligned} \sigma_n^2(\theta) &= \frac{1}{n} E_p(Y_n' S_n'(\theta) R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) Y_n), \\ &= \frac{\sigma^2}{n} \text{tr}(R_n S_n^{-1} S_n'(\theta) R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n). \end{aligned}$$

Hence,

$$Q_{p,n}(\theta) = -\frac{n}{2} \log(2\pi + 1) + \frac{n}{2} \log \sigma_n^2(\theta) - \log |R_n(\lambda)| + \log |S_n(\theta)|.$$

By Shannon-Kolmogorov Information Inequality (Ferguson (1996), p113),  $Q_{p,n}(\theta) \leq Q_{p,n}(\theta_0)$  for all  $\theta \in \Theta$ . This implies that  $\frac{1}{n}(Q_{p,n}(\theta) - Q_{p,n}(\theta_0)) \leq 0$  for all  $\theta \in \Theta$ .

**Proof of (iii)** We show that the identification uniqueness condition holds by contradiction.

$$\begin{aligned} \frac{1}{n}(Q_n(\theta) - Q_n(\theta_0)) &= -\frac{1}{2} \log \sigma_n^{*2}(\theta) - \log |R_n(\lambda)| + \log |S_n(\theta)| - \left( -\frac{1}{2} \log \sigma_0^2 - \log |R_n| + \log |S_n| \right) \\ &= \left( -\frac{1}{2} (\log \sigma_n^2(\theta) - \log \sigma_0^2) - \frac{1}{n} (\log |R_n(\lambda)| - \log |R_n|) + \right. \\ &\quad \left. \frac{1}{n} (\log |S_n(\theta)| - \log |S_n|) \right) - \frac{1}{n} (\log \sigma_n^{*2}(\theta) - \log \sigma_n^2(\theta)), \\ &= \frac{1}{n} (Q_{p,n}(\theta) - Q_{p,n}(\theta_0)) - \frac{1}{2} (\log \sigma_n^{*2}(\theta) - \log \sigma_n^2(\theta)). \end{aligned}$$

Moreover,

$$\sigma_n^{*2}(\theta) - \sigma_n^2(\theta) = \frac{1}{n} \beta'_0 X'_n S_n'^{-1} S'_n(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0.$$

$M_n$  is positive semi definite and thereby  $\sigma_n^{*2}(\theta) - \sigma_n^2(\theta) \geq 0$ . This implies  $-\frac{1}{2} (\log \sigma_n^{*2}(\theta) - \log \sigma_n^2(\theta)) \leq 0$ .

Now, suppose that the identification uniqueness condition does not hold. Then, there exists an  $\epsilon > 0$  and a sequence  $\{\theta_n\}$  in  $N_\epsilon^c(\theta_0)$  such that  $\lim_{n \rightarrow \infty} \frac{1}{n} (Q_n(\theta) - Q_n(\theta_0)) = 0$ . By the compactness of  $N_\epsilon^c(\theta_0)$ , there exists a convergent subsequence  $\{\theta_{n_m}\}$  of  $\{\theta_n\}$  with the limit  $\theta_+$  of  $\theta_{n_m}$  being in  $N_\epsilon^c(\theta_0)$ . This implies that  $\theta_+ \neq \theta_0$ . As  $\frac{1}{n} Q_n(\theta)$  is uniformly equicontinuous,  $\lim_{n_m \rightarrow \infty} \frac{1}{n_m} (Q_{n_m}(\theta_+) - Q_{n_m}(\theta_0)) = 0$ . Because  $\frac{1}{n} (Q_{p,n}(\theta) - Q_{p,n}(\theta_0)) \leq 0$  and  $-\frac{1}{2} (\log \sigma_n^{*2}(\theta) - \log \sigma_n^2(\theta)) \leq 0$ , this is possible only if  $\lim_{n_m \rightarrow \infty} \frac{1}{n_m} (Q_{n_m}(\theta_+) - Q_{n_m}(\theta_0)) = 0$  and  $-\frac{1}{2} (\log \sigma_n^{*2}(\theta) - \log \sigma_n^2(\theta)) \leq 0$ . However,  $\lim_{n \rightarrow \infty} \frac{1}{n} \beta'_0 X'_n S_n'^{-1} S'_n(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0 \neq 0$  from the assumption in Theorem 3. Thus,  $-\frac{1}{2} (\log \sigma_n^{*2}(\theta) - \log \sigma_n^2(\theta)) < 0$  and consequently

$\lim_{n_m \rightarrow \infty} \frac{1}{n_m} (Q_{n_m}(\theta_+) - Q_{n_m}(\theta_0)) \neq 0$ . This is a contradiction. Therefore, the identification uniqueness condition must hold.

The consistency of  $\hat{\theta}$  follow from uniform convergence and the identification uniqueness condition. This completes the proof of the theorem.  $\square$

## Proof of Theorem 4

We have by the Taylor expansion,

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\hat{\psi}_n)}{\partial \psi}, \\ &= \frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \psi} + \left( \frac{1}{n} \frac{\partial^2 \log L_n(\bar{\psi}_n)}{\partial \psi \partial \psi'} \right) \sqrt{n} (\hat{\psi}_n - \psi_0), \end{aligned}$$

where  $\bar{\psi}_n$  lies between  $\hat{\psi}_n$  and  $\psi_0$ . Thus, the asymptotic normality of  $\hat{\psi}_n$  follows if

- (a)  $\frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \psi} \xrightarrow{D} N(0, \lim_{n \rightarrow \infty} \Gamma(\psi_0)),$
- (b)  $\frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} - E\left(\frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'}\right) \xrightarrow{p} 0,$  and
- (c)  $\frac{1}{n} \frac{\partial^2 \log L_n(\bar{\psi}_n)}{\partial \psi \partial \psi'} - \frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} \xrightarrow{p} 0.$

**Proof of (a)** The asymptotic normality of  $\frac{1}{\sqrt{n}} \frac{\partial \log L_n(\psi_0)}{\partial \psi}$  follows from the central limit theorems for linear-quadratic forms in Kelejian and Prucha (2001). We need to check that the score vector holds Assumptions in Kelejian and Prucha (2001). To check assumptions for asymptotic normality, it is sufficient to show some matrices hold desired boundaly conditions. From assumptions of this paper and Lemma 3.5.2,  $(R'_n S_n'^{-1} W'_n R_n'^{-1} - W'_n R_n'^{-1})$  and  $R'_n S_n'^{-1} W'_n R_n'^{-1}$  are uniformly bounded in column sums, and the elements of  $X'_n S_n'^{-1} W'_n R_n'^{-1}$  are uniformly bounded. Thus, each score function holds the assumptions and the asymptotic normality of each score function follows. Finally, the Cramér-Wold devise (Proposition 6.3.1 of Brockwell and Davis (1991)) leads to the joint asymptotic normality.

**Proof of (b)** Let  $D_{\psi\psi}$  be  $\frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} - E\left(\frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'}\right)$ . Then,  $D_{\psi\psi}$  has the elements:

$$\begin{aligned}
D_{\beta\beta'} &= 0, \\
D_{\beta\sigma^2} &= -\frac{1}{n\sigma_0^4} X'_n R_n'^{-1} V_n, \\
D_{\beta\rho} &= -\frac{1}{n\sigma_0^2} X'_n R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n V_n, \\
D_{\beta\lambda} &= \frac{1}{n\sigma_0^2} X'_n (R_n'^{-1} W_n' R_n'^{-1} + R_n'^{-1} R_n^{-1} W - R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n) V_n, \\
D_{\sigma^2\sigma^2} &= \frac{1}{\sigma_0^4} - \frac{1}{n\sigma_0^6} V_n' V_n, \\
D_{\sigma^2\rho} &= -\frac{1}{n\sigma_0^4} \beta'_0 X'_n S_n'^{-1} W_n' R_n'^{-1} V_n - \frac{1}{n\sigma_0^4} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} V_n - \sigma_0^2 \text{tr}(S_n'^{-1} W_n')), \\
D_{\sigma^2\lambda} &= -\frac{1}{n\sigma_0^4} \beta'_0 X'_n S_n'^{-1} W_n' R_n'^{-1} V_n + \frac{1}{n\sigma_0^4} (V_n' W_n' R_n'^{-1} V_n - \sigma_0^2 \text{tr}(W_n' R_n'^{-1})) \\
&\quad - \frac{1}{n\sigma_0^4} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} V_n - \sigma_0^2 \text{tr}(S_n'^{-1} W_n')), \\
D_{\rho\rho} &= -\frac{2}{n\sigma_0} \beta'_0 X'_n S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n V_n \\
&\quad - \frac{1}{n\sigma_0^2} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n V_n - \sigma_0^2 \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n)), \\
D_{\rho\lambda} &= \frac{1}{n\sigma_0^2} \beta'_0 X'_n (S_n'^{-1} W_n' R_n'^{-1} W_n' R_n'^{-1} + S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n - 2S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n) V_n \\
&\quad + \frac{1}{n\sigma_0^2} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} W_n' R_n'^{-1} V_n - \sigma_0^2 \text{tr}(S_n'^{-1} W_n' R_n'^{-1} W_n')) \\
&\quad + \frac{1}{n\sigma_0^2} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n V_n - \sigma_0^2 \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n)) \\
&\quad - \frac{1}{n\sigma_0^2} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n V_n - \sigma_0^2 \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n)), \\
D_{\lambda\lambda} &= \frac{1}{n\sigma_0^2} \beta'_0 X'_n (2S_n'^{-1} W_n' R_n'^{-1} W_n' R_n'^{-1} + S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n - 2S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n'^{-1} R_n \\
&\quad - 2R_n'^{-1} W_n' R_n'^{-1} W_n' R_n'^{-1} - R_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n + 2R_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n \\
&\quad + 2R_n'^{-1} W_n' R_n'^{-1} W_n' R_n'^{-1} + R_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n - R_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n) V_n \\
&\quad + \frac{2}{n\sigma_0^2} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} W_n' R_n'^{-1} V_n - \sigma_0^2 \text{tr}(S_n'^{-1} W_n' R_n'^{-1} W_n')) \\
&\quad + \frac{1}{n\sigma_0^2} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n V_n - \sigma_0^2 \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n)) \\
&\quad - \frac{1}{n\sigma_0^2} (V_n' R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n V_n - \sigma_0^2 \text{tr}(R_n' S_n'^{-1} W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n)) \\
&\quad - \frac{2}{n\sigma_0^2} (V_n' W_n' R_n'^{-1} W_n' R_n'^{-1} V_n - \sigma_0^2 \text{tr}(W_n' R_n'^{-1} W_n' R_n'^{-1})) \\
&\quad + \frac{1}{n\sigma_0^2} (V_n' W_n' R_n'^{-1} R_n^{-1} W_n V_n - \sigma_0^2 \text{tr}(W_n' R_n'^{-1} R_n^{-1} W_n)) \\
&\quad + \frac{1}{n\sigma_0^2} (V_n' W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n V_n - \sigma_0^2 \text{tr}(W_n' R_n'^{-1} R_n^{-1} W_n S_n^{-1} R_n)).
\end{aligned}$$



Thus, the elements of  $D_{\psi\psi}$  are decomposed into sums of the forms:  $\frac{1}{n}X'_n A_n(\theta)V_n$ ,  $\frac{1}{n}\beta'_0 X'_n A_n(\theta)V_n$ ,  $\frac{1}{n}(V'_n A_n(\theta)V_n - E(V'_n A_n(\theta)V_n))$  and  $\frac{1}{\sigma_0^4} - \frac{1}{n\sigma_0^6}V'_n V_n$ , where a matrix  $A_n(\theta)$  is uniformly bounded in both row and column sums. From Lemma 3.5.3,  $\frac{1}{n}X'_n A_n(\theta)V_n$ ,  $\frac{1}{n}\beta'_0 X'_n A_n(\theta)V_n$  and  $\frac{1}{n}(V'_n A_n(\theta)V_n - E(V'_n A_n(\theta)V_n))$  are convergence to zero in probability. Moreover,  $\frac{1}{\sigma_0^4} - \frac{1}{n\sigma_0^6}V'_n V_n \xrightarrow{P} 0$  because  $\frac{1}{n}V'_n V_n \xrightarrow{P} \sigma_0^2$  by the law of large numbers. Therefore, it follow that  $\frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} - E(\frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'}) \xrightarrow{P} 0$ .

**Proof of (c)** From Lemma 3.5.2 and 3.5.3, it is easy to show that  $\frac{1}{n} \frac{\partial^2 \log L_n(\bar{\psi}_n)}{\partial \psi \partial \psi'} = O_p(1)$  and  $\frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} = O_p(1)$ . Here,  $\bar{\sigma}^{-r} = \sigma_0^{-r} + o_p(1)$ ,  $r = 2, 4, 6$  because  $\bar{\sigma}^2 \xrightarrow{P} \sigma_0^2$  and  $\sigma^r$  appears in  $H_n(\psi) \equiv \frac{\partial^2}{\partial \psi \partial \psi'} \log L_n(\psi)$  multiplicatively, thus it results in an asymptotically negligible error to replace  $\bar{\sigma}^2$  by  $\sigma_0^2$ . The elements of the Hessian matrix,  $H_n(\psi) \equiv \frac{\partial^2}{\partial \psi \partial \psi'} \log L_n(\psi)$ , are decomposed into sums of terms of the forms:  $X'_n A_n(\theta)X_n$ ,  $X'_n A_n(\theta)Y_n$ ,  $X'_n A_n(\theta)V(\theta)$ ,  $Y'_n A_n(\theta)Y_n$ ,  $\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}V'_n(\theta)V_n(\theta)$ ,  $Y'_n A_n(\theta)V_n(\theta)$ ,  $V'_n(\theta)A_n(\theta)V_n(\theta)$  and  $tr(A_n(\theta))$ , where a matrix  $A_n(\theta)$  is uniformly bounded in both row and column sums. Therefore, it is sufficient to show that the difference between each term at  $\bar{\psi}$  and  $\psi_0$  converges to zero in probability and moreover this can be easily shown. We show some examples corresponding each term of the Hessian matrix.

Noting that

$$\begin{aligned} R_n^{-1}(\lambda) - R_n^{-1} &= R_n^{-1}(\lambda)(R_n - R_n(\lambda))R_n^{-1}, \\ &= (\lambda_0 - \lambda)R_n^{-1}(\lambda)W_n R_n^{-1}. \end{aligned}$$

For  $X'_n A_n(\theta)X_n$ ,

$$\begin{aligned} \frac{1}{n}X'_n R_n^{-1}(\bar{\lambda})R_n^{-1}(\bar{\lambda})X_n - \frac{1}{n}X'_n R_n^{-1}R_n^{-1}X_n &= \frac{1}{n}X'_n(R_n^{-1}(\bar{\lambda}) - R_n'^{-1} + R_n'^{-1})R_n^{-1}(\bar{\lambda})X_n - \\ &\quad \frac{1}{n}X'_n R_n'^{-1}R_n^{-1}X_n, \\ &= \frac{1}{n}X'_n(R_n'^{-1}(\bar{\lambda}) - R_n'^{-1})R_n^{-1}(\bar{\lambda})X_n + \\ &\quad \frac{1}{n}X'_n R_n'^{-1}R_n^{-1}(\bar{\lambda})X_n - \frac{1}{n}X'_n R_n'^{-1}R_n^{-1}X_n, \\ &= (\lambda_0 - \bar{\lambda})\frac{1}{n}X'_n R_n'^{-1}(\lambda)W'_n R_n^{-1}R_n^{-1}(\bar{\lambda})X_n \\ &\quad + (\lambda_0 - \bar{\lambda})\frac{1}{n}X'_n R_n'^{-1}R_n^{-1}(\lambda)W_n R_n^{-1}X_n, \\ &= o_p(1)O(1) + o_p(1)O(1), \\ &= o_p(1). \end{aligned}$$

Moreover, the convergence of  $X'_n A_n(\theta)Y_n$  is shown similarly.

Noting that

$$\begin{aligned}
V_n(\theta) &= R_n^{-1}(\lambda)R_n(\lambda)V_n(\theta), \\
&= R_n^{-1}(\lambda)(S(\theta)Y_n - X_n\beta), \\
&= R_n^{-1}(\lambda)((\lambda_0 - \lambda)W_nY_n + (\rho_0 - \rho)W_nY_n + X_n(\beta_0 - \beta) + R_nV_n).
\end{aligned}$$

Thus, for  $X'_nA_n(\theta)V(\theta)$ ,

$$\begin{aligned}
\frac{1}{n}X'_nR_n'^{-1}(\bar{\lambda})V_n(\bar{\theta}) - \frac{1}{n}X'_nR_n'^{-1}V_n &= ((\lambda_0 - \bar{\lambda}) + (\rho_0 - \bar{\rho}))\frac{1}{n}X'_nR_n'^{-1}(\bar{\lambda})W_nY_n + \\
&\quad \frac{1}{n}X'_nR_n'^{-1}(\bar{\lambda})X_n(\beta_0 - \beta) \\
&\quad + \frac{1}{n}X'_nR_n'^{-1}(\bar{\lambda})R_nV_n - \frac{1}{n}X'_nR_n'^{-1}V_n, \\
&= o_p(1)O_p(1) + O_p(1)o_p(1) + o_p(1) + o_p(1), \\
&= o_p(1),
\end{aligned}$$

where the convergence of last two terms follow from Lemma 3.5.3.

Here,

$$\begin{aligned}
\frac{1}{n}V_n'(\bar{\theta})V_n(\bar{\theta}) &= ((\lambda_0 - \bar{\lambda}) + (\rho_0 - \bar{\rho}))^2\frac{1}{n}Y_n'W_n'R_n'^{-1}(\bar{\lambda})R_n^{-1}(\bar{\lambda})W_nY_n \\
&\quad + (\beta_0 - \beta)' \frac{1}{n}X'_nR_n'^{-1}(\bar{\lambda})R_n^{-1}(\bar{\lambda})X_n(\beta_0 - \beta) + \frac{1}{n}V_n'R_n'R_n'^{-1}(\bar{\lambda})R_n^{-1}(\bar{\lambda})R_nV_n \\
&\quad + \frac{2}{n}((\lambda_0 - \bar{\lambda}) + (\rho_0 - \bar{\rho}))Y_n'W_n'R_n'^{-1}(\bar{\lambda})R_n^{-1}(\bar{\lambda})X_n(\beta_0 - \beta) \\
&\quad + \frac{2}{n}((\lambda_0 - \bar{\lambda}) + (\rho_0 - \bar{\rho}))Y_n'W_n'R_n'^{-1}(\bar{\lambda})R_n^{-1}(\bar{\lambda})R_nV_n + \\
&\quad (\beta_0 - \beta)' \frac{2}{n}X'_nR_n'^{-1}(\bar{\lambda})R_n^{-1}(\bar{\lambda})R_nV_n, \\
&= o_p(1)O_p(1) + o_p(1)O(1)o_p(1) + \sigma_0^2 + o_p(1)O_p(1)o_p(1) + o_p(1)O_p(1) + o_p(1)o_p(1), \\
&= \sigma_0^2 + o_p(1).
\end{aligned}$$

It follows that  $\frac{1}{2\sigma_0^4} - \frac{1}{n\sigma_0^6}V_n'(\theta)V_n(\theta) = o_p(1)$ .

Before next proof, we show an example.  $Y_n'S_n(\theta)V_n = \beta'X_n'S_n^{-1}S(\theta)V_n + V_n'R_n'S_n^{-1}S_n(\theta)V_n$  and

$$\begin{aligned}
\frac{1}{n}V_n'R_n'S_n^{-1}S_n(\theta)V_n - \frac{1}{n}V_n'R_n'S_n^{-1}S_nV_n &= ((\lambda_0 - \lambda) + (\rho_0 - \rho))\frac{1}{n}V_n'R_n'S_n^{-1}V_n, \\
&= o_p(1)O_p(1), \\
&= o_p(1).
\end{aligned}$$

It follows that  $\frac{1}{n}Y_n'S_n(\theta)V_n - \frac{1}{n}Y_n'S_nV_n = o_p(1)$  and similarly  $\frac{1}{n}Y_n'A_n(\theta)V_n - \frac{1}{n}Y_n'A_nV_n = o_p(1)$  and  $\frac{1}{n}Y_n'A_n(\theta)Y_n - \frac{1}{n}Y_n'A_nY_n = o_p(1)$  where  $A_n$  is  $A_n(\theta)$  at true value  $\theta_0$ .

Now, for  $Y_n' A_n(\theta) V_n(\theta)$ ,

$$\begin{aligned}
\frac{1}{n} Y_n' W_n' R_n'^{-1}(\lambda) V_n(\theta) - \frac{1}{n} Y_n' W_n' R_n'^{-1} V_n &= ((\lambda_0 - \bar{\lambda}) + (\rho_0 - \bar{\rho})) \frac{1}{n} Y_n' W_n' R_n'^{-1}(\lambda) R_n^{-1}(\lambda) W_n Y_n \\
&\quad + \frac{1}{n} Y_n' W_n' R_n'^{-1}(\lambda) R_n^{-1}(\lambda) X_n (\beta_0 - \bar{\beta}) \\
&\quad + \frac{1}{n} Y_n' W_n' R_n'^{-1}(\bar{\lambda}) R_n^{-1}(\bar{\lambda}) R_n V - \frac{1}{n} Y_n' W_n' R_n'^{-1} V_n \\
&= o_p(1) O_p(1) + O_p(1) o_p(1) + o_p(1) \\
&= o_p(1).
\end{aligned}$$

Moreover, the convergence of  $V_n(\theta)' A_n(\theta) V_n(\theta)$  is also shown similarly.

Finally, for  $tr(A_n(\theta))$ , by the Taylor expansion,

$$\begin{aligned}
\frac{1}{n} tr(R_n^{-1}(\lambda) W_n R_n^{-1}(\lambda) W_n) - \frac{1}{n} tr(R_n^{-1} W_n R_n^{-1} W_n) &= \frac{d}{d\lambda} tr(R_n^{-1}(\tilde{\lambda}) W_n R_n^{-1}(\tilde{\lambda}) W_n) (\bar{\lambda} - \lambda_0), \\
&= O(1) o_p(1), \\
&= o_p(1),
\end{aligned}$$

where  $\tilde{\lambda}$  lies between  $\bar{\lambda}$  and  $\lambda_0$ .

The convergence of the other elements of the Hessian matrix are shown similarly, hence  $\frac{1}{n} \frac{\partial^2 \log L_n(\bar{\psi}_n)}{\partial \psi \partial \psi'} - \frac{1}{n} \frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} \xrightarrow{P} 0$ .

This completes the proof of the theorem.  $\square$

## Proof of Theorem 5

The estimator for  $\alpha$  is

$$\hat{\alpha}_n = (1 - \hat{\lambda}) \log \left( \frac{1}{n} \sum_{i=1}^n \exp\{(R_n^{-1}(\hat{\lambda})[S(\hat{\theta})Y_n - Z_n \hat{\delta}])_i\} \right),$$

Here,

$$\begin{aligned}
S(\hat{\theta})Y_n - Z_n \hat{\delta} &= Y_n - \hat{\lambda} W_n Y_n - \hat{\rho} W_n Y_n - Z_n \hat{\delta}, \\
&= (\lambda_0 - \hat{\lambda}) W_n Y_n + (\rho_0 - \hat{\rho}) W_n Y_n + Z_n (\delta_0 - \hat{\delta}) + \alpha_0 \mathbf{1}_n + R_n V_n, \\
&= D + \alpha_0 \mathbf{1}_n + R_n V_n,
\end{aligned}$$

where  $D = (\lambda_0 - \hat{\lambda}) W_n Y_n + (\rho_0 - \hat{\rho}) W_n Y_n + Z_n (\delta_0 - \hat{\delta})$ .

Because  $R_n^{-1}(\hat{\lambda})(S(\hat{\theta})Y_n - Z_n \hat{\delta}) = \frac{\alpha_0}{1-\hat{\lambda}} \mathbf{1}_n + R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n$ ,

$$\frac{1}{n} \sum_{i=1}^n \exp\{(R_n^{-1}(\hat{\lambda})[S(\hat{\theta})Y_n - Z_n \hat{\delta}])_i\} = \exp\left(\frac{\alpha}{1-\hat{\lambda}}\right) \frac{1}{n} \sum_{i=1}^n \exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n)_i\}.$$

Thus,

$$\hat{\alpha} - \alpha_0 = (1 - \hat{\lambda}) \log \left( \frac{1}{n} \sum_{i=1}^n \exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n)_i\} \right). \quad (3.10)$$

To prove consistency, it is sufficient that the right side of (3.10) converges to zero in probability.

By the Taylor expansion,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n)_i\} &= 1 + \frac{1}{n} \sum_{i=1}^n \exp(b_i) \{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n)_i\} \\ &= 1 + \frac{1}{n} \mathbf{b}'(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n), \end{aligned}$$

where  $b_i$  lies between 0 and  $(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n)_i$ , and  $\mathbf{b} = (b_1, \dots, b_n)'$ .

From Assumptions, Theorem 3 and Lemma 3.5.2 and 3.5.3,

$$\begin{aligned} \frac{1}{n} \mathbf{b}'(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n) &= (\lambda_0 - \hat{\lambda}) \frac{1}{n} \mathbf{b}' R_n^{-1}(\hat{\lambda}) W_n Y_n + (\rho_0 - \hat{\rho}) \frac{1}{n} \mathbf{b}' R_n^{-1}(\hat{\lambda}) W_n Y_n \\ &\quad + \frac{1}{n} \mathbf{b}' R_n^{-1}(\hat{\lambda}) Z_n (\delta_0 - \hat{\delta}) + \frac{1}{n} \mathbf{b}' R_n^{-1}(\hat{\lambda}) R_n V_n, \\ &= o_p(1) O_p(1) + o_p(1) O_p(1) + O(1) o_p(1) + o_p(1), \\ &= o_p(1). \end{aligned}$$

Thus,  $\frac{1}{n} \sum_{i=1}^n \exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n)_i\} \xrightarrow{p} 1$  and  $(1 - \hat{\lambda}) \log\left(\frac{1}{n} \sum_{i=1}^n \exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_n V_n)_i\}\right) \xrightarrow{p} 0$ .  $\square$

# Bibliography

- [1] Andrews, D. W. (1992). Generic uniform convergence. *Econometric theory*, 8(2), 241-257.
- [2] Bernstein, D. S. (2009). *Matrix mathematics: Theory, facts, and formulas*. 2nd edition. Princeton: Princeton University Press.
- [3] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
- [4] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics*, 498-505.
- [5] Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy*, 96(1), 116-131.
- [6] Borovkova, S. & Lopuhaa, R., (2012). Spatial GARCH: A Spatial Approach to Multivariate Volatility Modeling. Available at SSRN: <https://ssrn.com/abstract=2176781>.
- [7] Brockwell, P. J., & Davis, R. A. (1991). *Time series: theory and methods*. Springer-Verlag.
- [8] Caporin, M., and Paruolo, P., 2009, Structured multivariate volatility models, Available at SSRN: <http://ssrn.com/abstract=1318639>.
- [9] Davidson, J. (1994). *Stochastic limit theory*. Oxford: Oxford University Press.
- [10] Doan, O., & Tapnar, S. (2013). GMM estimation of spatial autoregressive models with moving average disturbances. *Regional Science and Urban Economics*, 43(6), 903-926.
- [11] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.

- [12] Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory*, 11(1), 122-150.
- [13] Fan, J., & Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.
- [14] Ferguson, T. S. (1996). *A course in large sample theory*. Chapman & Hall.
- [15] Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. New York: Springer.
- [16] Kelejian, H. H., & Prucha, I. R. (1997). Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. *International regional science review*, 20(1-2), 103-111.
- [17] Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99-121.
- [18] Kelejian, H. H., & Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2), 219-257.
- [19] Kelejian, H. H., & Robinson, D. P. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in regional science*, 72(3), 297-312.
- [20] Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric theory*, 18(2), 252-277.
- [21] Lee, L. F. (2004). Asymptotic Distributions of Quasi- Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica*, 72(6), 1899-1925.
- [22] Lee, L. F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137(2), 489-514.
- [23] Lee, P. M. (2012). *Bayesian statistics: an introduction*. John Wiley & Sons.
- [24] Robinson, P. M. (2009). Large- sample inference on spatial dependence. *The Econometrics Journal*, 12(s1).
- [25] Sato, T. and Matsuda, Y. (2017). Spatial Autoregressive Conditional Heteroskedasticity Models. *J.Japan Statist. Soc.*, Vol. 47, 2
- [26] Schott, J. R. (2005). *Matrix analysis for statistics*. 2nd edition. John Wiley & Sons.

- [27] Su, L., & Yang, Z. (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics*, 185(1), 230-258.
- [28] White, H. (1994). *Estimation, inference and specification analysis*. New York: Cambridge university press.
- [29] Yan, J. (2007). Spatial stochastic volatility for lattice data. *Journal of agricultural, biological, and environmental statistics*, 12(1), 25.
- [30] Yang Z. (2018). Unified M-Estimation of Fixed-Effects Spatial Dynamic Models with Short Panels. *Journal of Econometrics*, 205(2), 423-427.
- [31] Yu, J., De Jong, R., & Lee, L. F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large. *Journal of Econometrics*, 146(1), 118-134.

## Chapter 4

# SARAR-GARCH models

### Abstract

This study proposes spatio-temporal extensions of time series generalized autoregressive conditional heteroskedasticity (GARCH) models. We call spatio-temporally extended GARCH models as spatial autoregressive models with spatial autoregressive error and generalized autoregressive conditional heteroskedasticity processes, namely SARAR-GARCH models. One important problem which multivariate volatility models contain is the curse of dimensionality. To overcome the problem, we adopt a spatial weight matrix which expresses the dependence relation between observations. A spatial weight matrix is usually determined by geographical information of spatial data. However, financial data doesn't include geographical information. Therefore, we propose a method to make spatial weight matrix from financial data by stepwise backward regressions. Parameters are estimated by a two step procedure. First step is the estimation of spatial parameters and second step is that of GARCH parameters. In real data analysis, We apply the SARAR-GARCH model to daily returns of the Nikkei 225 stock price data and S&P 500 stock price data. We compare the in-sample and out-sample performances of SARAR-GARCH models with those of CCC models which is a benchmark. The results show the in-sample performance of the CCC model is better because the CCC model contains many more parameters. However, the out-sample performance of the SARAR-GARCH model are better than that of the CCC model in both markets analysis.

### 4.1 Introduction

Volatility which is a conditional variance in a model is one of the most important concepts in financial econometrics because it is used in widely areas such as risk management, option pricing and portfolio selection. Financial market data often exhibits volatility clustering (i.e., volatility may be high for certain time periods



and low for other periods) This means time-varying volatility is more common than constant volatility. Therefore, accurate modeling of time-varying volatility is important in financial econometrics.

The seminal work of Engle (1982) proposes autoregressive conditional heteroscedasticity (ARCH) models and the most important extension of the model is generalized ARCH (GARCH) models proposed by Bollerslev (1986). The models have been widely used to identify volatilities. After that, many extended GARCH models have been proposed. For example, integrated GARCH models (Engle and Bollerslev (1986)), exponential GARCH models (Nelson (1991)), threshold GARCH models (Glosten, et al (1993)), GARCH in the mean models, and GJR-GARCH models are proposed.

Univariate volatility models are generalized to multivariate cases in many ways. One important problem which multivariate volatility models contain is the curse of dimensionality. We estimate a conditional covariance matrix which has  $\frac{n(n+1)}{2}$  quantities for a  $n$ -dimensional time series, therefore it is difficult to estimate all quantities. Thus, we attempt to give a conditional covariance matrix some simple structures to reduce the number of parameters. For example, exponentially weighted moving average models, constant conditional correlation models (Bollerslev (1990)), BEKK models (Engle and Kroner (1995)), orthogonal GARCH models (Alexander (2001)), dynamic conditional correlation models (Tse and Tsui (2002)), dynamic orthogonal component models, and factor GARCH models are proposed.

The ideas of spatial econometrics have been applied to volatility models to reduce number of parameters in a covariance matrix in recent years. Caporin and Paruolo (2008) and Borovkova and Lopuhaa (2012) have applied the ideas of spatial econometrics to time series multivariate GARCH models. Yan (2007) and Robinson (2009) have done spatial extensions of stochastic volatility models which are another kind of volatility models. Sato and Matsuda (2017, 2018) have extend time series GARCH models to spatial models.

This paper contributes to extend GARCH models to spatiotemporal models which we call spatial autoregressive models with spatial autoregressive error and generalized autoregressive conditional heteroskedasticity processes, namely SARAR-GARCH models by using spatial econometrics ideas. The model is characterized by a spatial weight matrix which express cross-section correlations between assets and used to reduce the number of parameters. A spatial weight matrix is usually determined by geographical information of spatial data. However, financial data doesn't include geographical information. Therefore, we propose a method to make spatial weight matrix from financial data. we apply the multiple linear regression model and stepwise backward regression to calculate spatial weights in spatial weight matrices. Parameters are estimated by a two step procedure. First step is the estimation of spatial parameters and second step is that of GARCH parameters. Spatial parameters are estimated in first step. We regard volatilities in the model as constant variance and we apply quasi-maximum likelihood method with the model. After that we apply GARCH models with residuals derived from first step in second step. In

real data analysis, We apply the SARAR-GARCH model to daily returns of the Nikkei 225 stock price data and S&P 500 stock price data. We compare the in-sample and out-sample performances of SARAR-GARCH models with those of CCC models. First, we check the in-sample performances based on log-likelihood. The results show the log-likelihood of the CCC model is grater than that of SARAR-GARCH. This means model fitting of the CCC model is better. One reason is that the number of parameters in CCC models is more than five times of those of SARAR-GARCH models. Secondly, we compare out-sample performances by using quasi-likelihood loss function. The result shows the quasi-likelihood loss function of SARAR-GARCH models are smaller than that of CCC models. Then, the out-sample performance of SARAR-GARCH models is better. One reason is the CCC model may be over-fitting and it cause lower forecasting performance. Moreover, SARAR-GARCH models have better prediction performance in U.S. market analysis because stock price in U.S. market are more volatile and proposed models can capture sharp fluctuations.

The rest of paper proceeds as follows. Section 4.2 introduces SARAR-GARCH models. The estimation procedures are described in section 4.3. Section 4.4 examines empirical properties of SARAR-GARCH models by applying the models to real data such as stock price in the Japanese and the U.S. market. Section 4.5 discusses some concluding remarks.

## 4.2 Models

Let  $r_{i,t}$  be returns of financial instruments. We shall define SARAR-GARCH models to describe volatilities of return series  $r_{i,t}$  by

$$\mathbf{r}_t = \lambda W \mathbf{r}_t + \mathbf{u}_t \quad (4.1)$$

$$\mathbf{u}_t = \rho W \mathbf{u}_t + \boldsymbol{\varepsilon}_t \quad (4.2)$$

$$\varepsilon_{i,t} = \sigma_{i,t} f_{i,t}, \quad (4.3)$$

$$f_{i,t} \sim i.i.d(0, 1),$$

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \sigma_{i,t-1}^2 + \beta_i \varepsilon_{i,t-1}^2,$$

where  $\mathbf{r}_t = (r_{1,t}, \dots, r_{n,t})$ ,  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})$ ,  $\sigma_{i,t}$  is volatility,  $f_{i,t}$  is an independent and identically distributed (i.i.d) random variable with mean zeros and variance 1. The matrix  $W$ , n by n matrix, is called a spatial weight matrix and pre-determined before analysis. For parameters  $(\rho, \lambda, \omega_i, \alpha_i, \beta_i)'$ ,  $\rho$  and  $\lambda$  describes spatial interactions of return series and  $\omega_i, \alpha_i$  and  $\beta_i$  are GARCH parameters. The positivity of  $\sigma_{i,t}^2$  is ensured by the following sufficient restrictions:  $\omega_i > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_i \geq 0$ , and the sum  $\alpha_i + \beta_i < 1$  for stationarity. It has been known that SARMA is guaranteed to exist when  $|\lambda| + |\rho| < 1$ .

Let us consider the volatility matrix for SARAR-GARCH models. From (4.3), the variance matrix of  $\boldsymbol{\varepsilon}_t$ ,  $\Gamma$ , is a diagonal matrix whose components are  $\sigma_{i,t}^2$ , that is,  $\Gamma = \text{diag}(\sigma_{1,t}, \dots, \sigma_{n,t})$ . From equations (4.1) and (4.2),

$$\mathbf{r}_t = (I - \lambda W)^{-1} (I - \rho W)^{-1} \boldsymbol{\varepsilon}_t.$$

Therefore, the volatility matrix for SARAR-GARCH models,  $\Sigma_t$ , are

$$\Sigma_t = (I - \lambda W)^{-1}(I - \rho W)^{-1}\Gamma_t(I - \rho W)'^{-1}(I - \lambda W)'^{-1}, \quad (4.4)$$

where  $I$  is a identity matrix. Volatility  $\sigma_{i,t}^2$  changes over time, so the volatility matrix for SARAR-GARCH models expresses time-varying volatility structures in financial instruments and can capture dynamic correlations between. Moreover, spatial weight matrix in (4.4) express cross-sectional correlation between observations and plays important role to reduce the number of parameters for cross-sectional correlation and to overcome the curse of dimensionality.

A spatial weight matrix is usually determined by geographical information of spatial data and predetermined such as first-order contiguity relation or inverse distance between observations. However,  $r_{i,t}$  is financial data and doesn't include geographical information. Therefore, we need to determine financial distances to make a spatial weight matrix. Some author have proposed spatial weight matrix based on financial distance calculated from financial statement data such as dividend yields or market capitalizations. Here, we propose a method to make spatial weight matrices from financial data by stepwise backward regression. we apply the multiple linear regression model:

$$r_{i,t} = \delta_0 + \sum_{j \neq i}^n \delta_j r_{j,t} + z_{i,t},$$

where  $z_{i,t}$  follows i.i.d normal distribution. Then, we obtain the least square estimates  $\delta_j$  and those t-values. After that we check the minimum t-values and If the value is smaller than a critical value, for example 1.96, then we remove observations which have minimum t-value. Next, we regress  $r_{i,t}$  on  $n-2$  observations and we repeat this procedure until the minimum t-value is grater than the critical value.

### 4.3 Estimation

We shall propose estimation of the parameters  $(\rho, \lambda, \omega_i, \alpha_i, \beta_i)'$  in SARAR-GARCH models. Parameters are estimated by a two step procedure. First step is the estimation of  $\lambda$  and  $\rho$  and second step is that of  $\omega_i, \alpha_i, \beta_i$ .

Now, let us derive quasi likelihood function by regarding  $f_{i,t}$ 's as Gaussian variables with mean zero and variance  $\sigma_{i,t}^2$ . Then, the likelihood function of SARAR-GARCH models is

$$\log L = T \log |I - \lambda W| + T \log |I - \rho W| + \sum_{t=1}^T \sum_{i=1}^n \left( -\frac{1}{2} \log 2\pi \sigma_{j,t}^2 - \frac{1}{2} \frac{\varepsilon_{i,t}^2}{\sigma_{i,t}^2} \right).$$

Here, the number of parameters are  $3n + 2$  and optimization of all parameters simultaneously is a difficult task, so we adopt a two step procedure to reduce the number of parameters.

Parameters  $\rho$  and  $\lambda$  are estimated in first step. The parameters are estimated by quasi-likelihood estimation method, but we regard  $\sigma_{i,t}^2$  as constant, so variance in the model is homosekedastic and doesn't change over time. Gaussian likelihood function for first step estimation is derived by regarding  $f_{i,t}$  as independent Gaussian variables with mean zero and variance  $\sigma^2$ . Then the log likelihood function is

$$\log L = T \log |I - \lambda W| + T \log |I - \rho W| + -\frac{nT}{2} \log 2\pi\sigma^2 \sum_{t=1}^T \sum_{i=1}^n \left( -\frac{1}{2} \frac{\varepsilon_{i,t}^2}{\sigma^2} \right). \quad (4.5)$$

The QML estimator  $\hat{\lambda}$  and  $\hat{\rho}$  maximizes the log likelihood function (4.5).

We move to estimation of GARCH parameters. We have already obtained estimate of spatial parameters,  $\lambda$  and  $\rho$ . The residuals are obtained by

$$\hat{\varepsilon}_t = (I - \hat{\rho}W)(I - \hat{\lambda}W)\mathbf{r}_t,$$

where  $\hat{\lambda}$  and  $\hat{\rho}$  are estimates of spatial parameters in first step. we apply the GARCH (1, 1) model to the residuals. Let  $f_{i,t}$  in (4.3) be Gaussian white noise with unit variance. Then  $\varepsilon_{i,t}$  is an GARCH (1, 1) process if

$$\begin{aligned} \varepsilon_{i,t} &= \sigma_{i,t}^2 f_{i,t} \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i u_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2. \end{aligned}$$

Here, we use residuals  $\hat{\varepsilon}_t$  in stead of  $\varepsilon_t$ . Then, the log likelihood function of GARCH models is given by

$$\log L = \sum_{t=2}^T \left\{ -\frac{1}{2} \log(2\pi\sigma_{i,t}^2) - \frac{\hat{\varepsilon}_{i,t}^2}{2\sigma_{i,t}^2} \right\},$$

where  $\sigma_{i,t}^2 = \beta_{i,0} + \beta_{i,1}u_{i,t-1}^2 + \beta_{i,2}\sigma_{i,t-1}^2$  can be evaluated recursively. Maximizing this with respect to  $\omega_i, \alpha_i$  and  $\beta_i$ , we have the estimators of GARCH parameters.

## 4.4 Real data analysis

We examine empirical properties of SARAR-GARCH models by applying daily return data of Japanese and U.S markets to demonstrate practical performances of volatilities and co-volatilities identified by SARAR-GARCH models. Moreover, we show prediction performance and dynamic spillover effect of shock.

### 4.4.1 Japanese market analysis

We apply the SARAR-GARCH model to daily returns of the Nikkei 225 stock price data, that is the returns  $(r_{i,t})$  are computed as  $100(\log P_t - \log P_{t-1})$ ,

where  $P_t$  is the closing price and  $t$  is the time index referring to trading day  $t$ . The sampling period starts on April 1st, 2002 and ends on July 4th, 2016 for a total of 3500 returns. Moreover, we sample data for prediction from July 5th, 2016 to December 30, 2016. The number of firms are 201. Spatial wight matrices are made in accordance with the manner written in section 4.2.

We adopt constant conditional correlation (CCC) models as a benchmark. Let  $\mathbf{r}_t = (r_{1,t}, \dots, r_{n,t})$  be a  $n$ -dimensional vector process. CCC models are represented by the following equations

$$\begin{aligned}\mathbf{r}_t &= \mathbf{\Sigma}_t^{\frac{1}{2}} \boldsymbol{\varepsilon}_t, \\ \mathbf{\Sigma}_t &= \text{diag}(\sigma_{1,t}^2, \dots, \sigma_{n,t}^2), \\ \sigma_{i,t} &= \omega_i + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2, i = 1, \dots, n\end{aligned}$$

where  $\mathbf{\Sigma}_t$  is a diagonal matrix with  $\sigma_{i,t}^2$  as  $i$ th diagonal element, and  $\boldsymbol{\varepsilon}_t$  unobservable random vector with mean equal to 0 and variance-covariance equal to  $\mathbf{R}_t = (\rho_{t,i,j})$ . CCC models assume the correlation matrix is constant.

Table 4.1: Estimated values of  $\lambda, \rho$  and GARCH parameters and their standard errors (s.e.) of  $\lambda$  and  $\rho$  in the SARAR-GARCH model applied to log returns of stock price data of Japanese financial market.

parameter	estimate	s.e
$\hat{\lambda}$	0.9200	0.0004
$\hat{\rho}$	-0.3428	0.0016
$\hat{\alpha}_i$	[0.02, 0.42]	
$\hat{\beta}_i$	[0.25, 0.98]	

Table 4.1 shows the estimated values of  $\lambda$  and  $\rho$ . Estimates of  $\alpha_i$  and  $\beta_i$  are in the ranges [0.02, 0.42] and [0.25, 0.98], respectively. We find that  $\hat{\lambda}$ , the strength of interactions among return series, are significant. This suggests that asset returns tend to move together strongly. Figure 4.1 shows the estimated volatilities of Mitsubishi UFJ financial group and Mizuho financial group and their co-volatilities. Japanese economy is boom in 2002 and The financial crisis occurs in 2008, namely  $T$  is around 1600, and co-volatilities of two companies are high. This means the connection of movement of stock prices of two companies become high in boom or depression period.

We compare the in-sample and out-sample performances of SARAR-GARCH models with those of CCC models. First, we check the in-sample performances based on log-likelihood. Table 4.2 shows the log-likelihood of CCC is bigger than that of SARAR-GARCH. This means model fitting of the CCC model is better. One reason is that the number of parameters in CCC models is more than five times of those of SARAR-GARCH models. Secondly, we compare out-sample performances. We calculate predicted volatility based on definition of the

Table 4.2: Log-likelihoods and quasi-likelihood loss functions for the SARAR-GARCH model and the CCC model applied to log returns of stock price data of Japanese financial market.

	in-sample log-likelihood	out-sample QLIKE
SARAR-GARCH	-177	279
CCC	-175	286

models. After that we calculate prediction error based on the quasi-likelihood loss function:

$$QLIKE = \frac{1}{T_{pre}} \sum_{t=1}^{T_{pre}} r_t' V_t^{-1} r_t + \log |V_t|,$$

where  $r_t$  is a vector of return series  $V_t$  is a volatility matrix made by predicted volatility and  $T_{pre}$  is the size of time dimension for prediction period. Table 4.2 shows out-sample performance of SARAR-GARCH models are better. This shows CCC models may be over-fitting and it cause lower forecasting performance. Small prediction errors are one advantage point of proposed models because predicted volatility plays an important role in risk management.

Figure 4.2 shows the spillover effect of shock. We assume only Mitsubishi UFJ financial group's return increase 1 percent and calculate the effect of this shock to other companies. Here, we choose three companies, namely, Sumitomo Mitusi financial group, Mizuho financial group and Sumitomo Mitusi real estate. The figure shows the effect to the companies in same sector is larger than the effect to other sectors and the effect converges to zero as time goes by.

#### 4.4.2 U.S. market analysis

We apply the SARAR-GARCH model to daily returns of the S&P 500 stock price data, that is the returns  $(r_{i,t})$ . The sampling period is same as that of Japanese market analysis case and the number of firms are 395. Moreover, spatial weight matrices are made in accordance with the manner written in section 4.2.

Table 4. 3 shows the estimated values of  $\lambda$  and  $\rho$ . Estimates of  $\alpha_i$  and  $\beta_i$  are in the ranges  $[0.01, 0.59]$  and  $[0.27, 0.98]$ , respectively.

We compare the in-sample and out-sample performances of SARAR-GARCH models with those of CCC models. First, we check the in-sample performances based on log-likelihood. Table 4.4 shows the log-likelihood of CCC is bigger than that of SARAR-GARCH. This means model fitting of the CCC model is better. One reason is that the number of parameters in CCC models is more than five times of those of SARAR-GARCH models. Secondly, we compare out-sample performances. Table 4.4 shows out-sample performance of SARAR-GARCH models are better. Moreover difference between QLIKE of SARAR-

Table 4.3: Estimated values of  $\lambda, \rho$  and GARCH parameters and their standard errors (s.e.) of  $\lambda$  and  $\rho$  in the SARAR-GARCH model applied to log returns of stock price data of U.S financial market.

parameter	estimate	s.e
$\hat{\lambda}$	0.9199	0.0006
$\hat{\rho}$	-0.3200	0.0017
$\hat{\alpha}_i$	[0.01, 0.59]	
$\hat{\beta}_i$	[0.27, 0.98]	

GARCH models and CCC models are larger in U.S market analysis. This shows SARAR-GARCH models work quite well in U.S market analysis. The reason why proposed models work well in U.S market is stock prices in U.S market are more volatile. CCC models assume constant correlation between stock prices so can't capture dynamic relations, but SARAR-GARCH models can capture dynamic correlation as volatility matrix for the model shown . Therefore, SARAR-GARCH models work well in U.S market.

Table 4.4: Log-likelihoods and quasi-likelihood loss functions for the SARAR-GARCH model and the CCC model applied to log returns of stock price data of U.S. financial market.

	in-sample log-likelihood	out-sample QLIKE
SARAR-GARCH	556	414
CCC	534	455

## 4.5 Conclusion

We have proposed a spatial autoregressive moving average models with generalized autoregressive conditional heteroskedasticity processes, namely SARAR-GARCH models to evaluate volatilities of financial instruments. We apply spatial weight matrices which is an important tool in spatial econometrics for multivariate volatility models to overcome the curse of dimensionality. we propose the two step procedure to estimate the parameters in SARAR-GARCH models. In the real data analysis of Japanese and U.S. markets, we detect SARAR-GARCH have smaller prediction error than that of CCC models.

We complete the paper by describing challenging problem for future research. In the empirical analysis, we used the spatial weight matrix based on least-squares estimates. However, The choice of spatial weight matrix is an important problem in spatial analysis for financial data. Therefore, another spatial weight matrix can be more interesting to improve our volatility analysis. Another challenge is to establish asymptotic properties of estimators for

SARAR-GARCH models to investigate theoretical properties of proposed estimators.



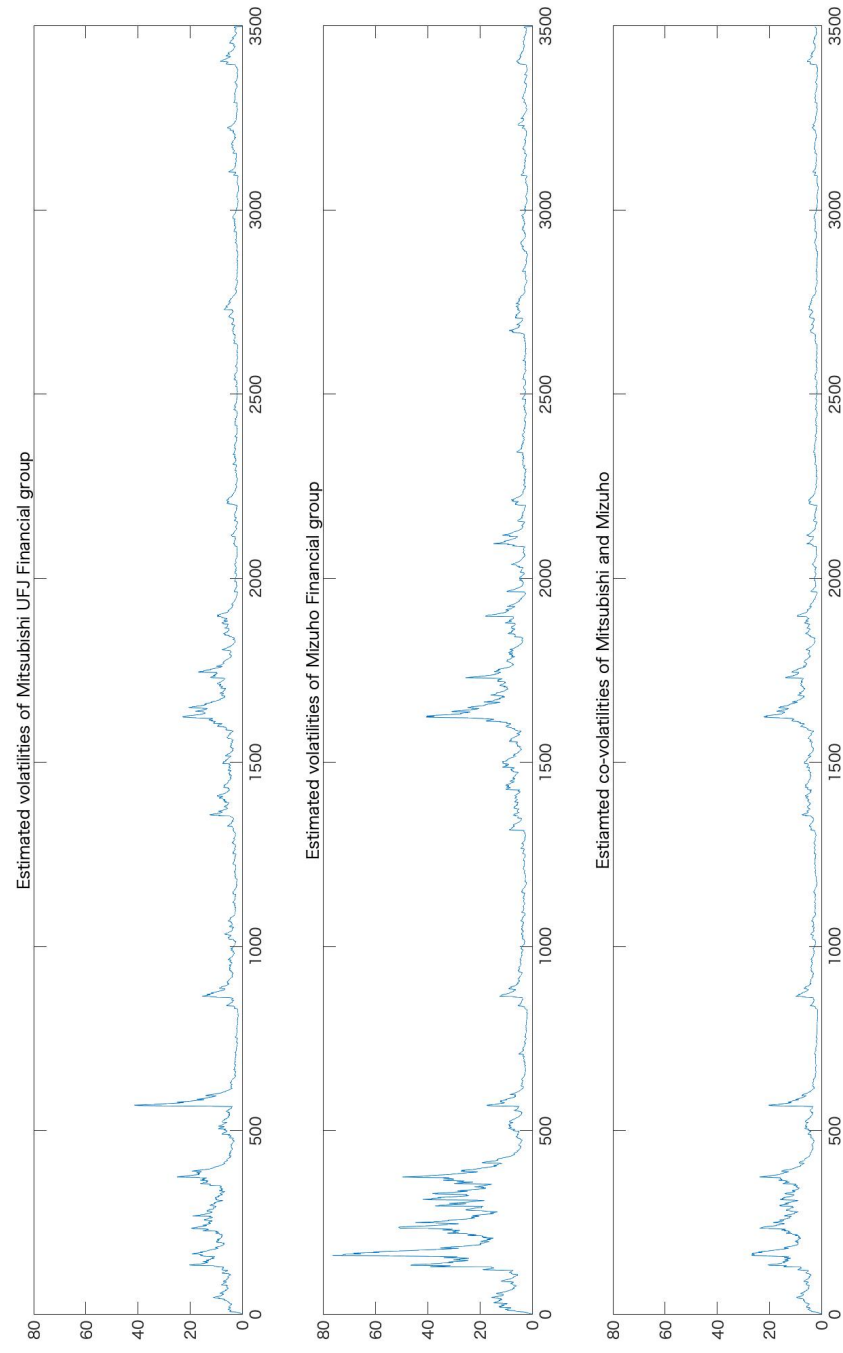


Figure 4.1: The identified volatilities of Mitsubishi UFJ financial group and Mizuho Financial group and their identified co-volatilities

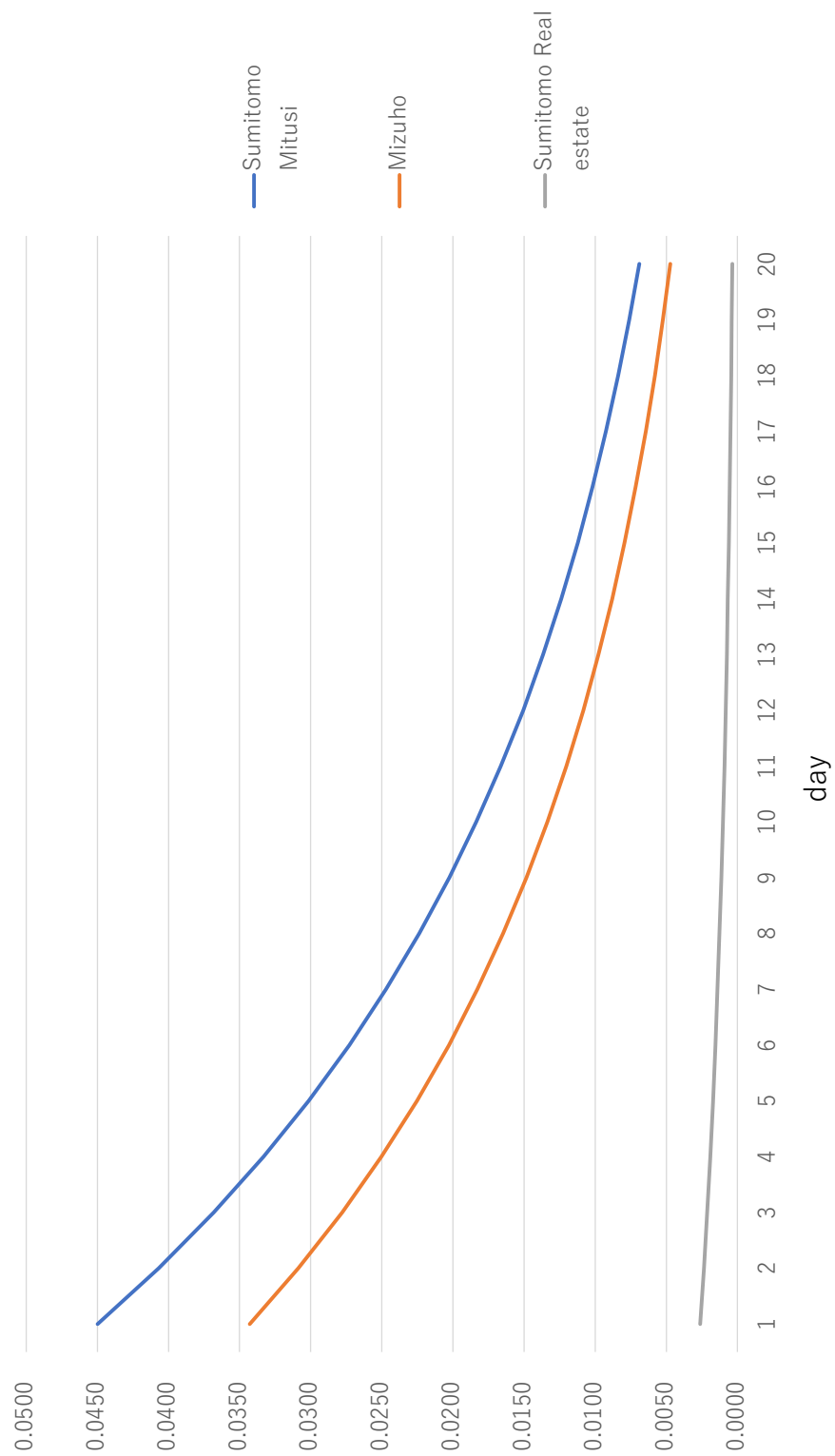


Figure 4.2: The identified spillover effect of the shock of Mitsubishi UFJ financial group to Sumitomo Mistui financial group, Misuho financial group and Sumitomo Real estate

# Bibliography

- [1] Alexander, C. (2001). Market models: A guide to financial data analysis. John Wiley & Sons.
- [2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
- [3] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics*, 498-505.
- [4] Borovkova, S. & Lopuhaa, R., (2012). Spatial GARCH: A Spatial Approach to Multivariate Volatility Modeling. Available at SSRN: <https://ssrn.com/abstract=2176781>.
- [5] Caporin, M., & Paruolo, P., 2008, Structured multivariate volatility models, Available at SSRN: <http://ssrn.com/abstract=1318639>.
- [6] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.
- [7] Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1), 1-50.
- [8] Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory*, 11(1), 122-150.
- [9] Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5), 1779-1801.
- [10] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.
- [11] Robinson, P. M. (2009). Large- sample inference on spatial dependence. *The Econometrics Journal*, 12(s1).
- [12] Sato, T. & Matsuda, Y. (2017). Spatial Autoregressive Conditional Heteroskedasticity Models. *J.Japan Statist. Soc.*, Vol. 47, 2

- [13] Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, 20(3), 351-362.
- [14] Yan, J. (2007). Spatial stochastic volatility for lattice data. *Journal of agricultural, biological, and environmental statistics*, 12(1), 25.